UNIVERSIDAD
DE MÁLAGA

Global Optimality via Tight Convex Relaxations for Pose Estimation in Geometric 3D Computer Vision

Autor: Jesus Briales Garcia

Director: Javier González Jiménez

Tesis doctoral
Doctorado en Ingeniería Mecatrónica
Dpt. de Ingeniería de Sistemas y Automática
Universidad de Málaga

2020

AUTOR: Jesús Briales García

http://orcid.org/0000-0002-4786-4936

EDITA: Publicaciones y Divulgación Científica. Universidad de Málaga

Prof. Dr. Javier González Jiménez is the supervisor of the doctoral dissertation entitled "Global Optimality via Tight Convex Relaxations for Pose Estimation in Geometric 3D Computer Vision" written by Jesus Briales Garcia. He hereby declares that this dissertation is suitable for the attainment of the degree "Doctor en Ingeniería Mecatrónica" awarded by the Universidad de Málaga.

Málaga, 14th December 2020

Prof. Dr. Javier González Jiménez

UNIVERSIDAD
DE MÁLAGA

ANDALUCÍA TECH
Campus de Excelencia Internacional

Escuela de Doctorado

DECLARACIÓN DE AUTORÍA Y ORIGINALIDAD DE LA TESIS PRESENTADA PARA OBTENER EL TÍTULO DE DOCTOR

D./Dña Jesús Briales García
Estudiante del programa de doctorado de Mecatrónica de la Universidad de Málaga, autor/a de la tesis, presentada para la obtención del título de doctor por la Universidad de Málaga, titulada: Global Optimality via Tight Convex Relaxations for Pose Estimation in Geometric 3D Computer Vision

Realizada bajo la tutorización y dirección de Javier González Jiménez (si tuviera varios directores deberá hacer constar el nombre de todos)

DECLARO QUE:

La tesis presentada es una obra original que no infringe los derechos de propiedad intelectual ni los derechos de propiedad industrial u otros, conforme al ordenamiento jurídico vigente (Real Decreto Legislativo 1/1996, de 12 de abril, por el que se aprueba el texto refundido de la Ley de Propiedad Intelectual, regularizando, aclarando y armonizando las disposiciones legales vigentes sobre la materia), modificado por la Ley 2/2019, de 1 de marzo.

Igualmente asumo, ante a la Universidad de Málaga y ante cualquier otra instancia, la responsabilidad que pudiera derivarse en caso de plagio de contenidos en la tesis presentada, conforme al ordenamiento jurídico vigente.

En Málaga, a 21 de Diciembre de 2020

Fdo.: Jesús Briales García

A mi familia.

Contents

Abstract

Artificial Intelligence (AI) is on the rise. It already drives a lot of services and products we use everyday. But for AI to bring its full potential into daily tasks and fuse seamlessly into our lives, with technologies such as autonomous driving, augmented reality or mobile robots, AI needs to be not only intelligent but also *perceptive*. In particular, the ability to *see* and to construct an accurate model of the environment is an essential capability to build intelligent perceptive systems.

Fortunately, research has not been idle in this direction. The ideas developed in Computer Vision for the last decades in areas such as Multiple View Geometry [1] or Optimization, put together to work into 3D reconstruction algorithms (be it Structure from Motion (SfM), Simultaneous Location and Mapping (SLAM), or Bundle Adjustment (BA)), seem to be mature enough to nurture a range of emerging applications that already employ as of today 3D Computer Vision in the background.

However, while there is a positive trend in the use of 3D reconstruction tools in real applications, there are also some fundamental limitations regarding reliability and performance guarantees that may hinder a wider adoption, e.g. in more critical applications involving people's safety such as autonomous navigation.

State-of-the-art 3D reconstruction algorithms typically formulate the reconstruction problem as a Maximum Likelihood Estimation (MLE) instance, which entails solving a high-dimensional non-convex non-linear optimization problem. In practice, this is done via fast local optimization methods that pursue the computation of a critical point of the MLE objective. Thus, having a good initialization is critical to reach the right critical point: The optimal solution. This bootstrapping is typically done by splitting the overall reconstruction complexity into intermediate, smaller subproblems, for which the literature has also given a wide battery of approaches and solutions. Yet, even for these simpler subproblems we often lack the ability to produce certifiably optimal solutions due to the different challenges surrounding non-convex optimization. Overall, this has driven to the current paradigm: The significant advances in 3D Computer Vision have enabled fast and scalable reconstruction pipelines, yet the lack of guarantees on most of the building blocks leave us with fundamentally brittle pipelines where small to no control in performance or guarantees exists.

One particularly ubiquitous scenario which is (maybe surprisingly) still hard to deal with is optimizing over 3D poses. A whole bunch of problems in 3D reconstruction, from the simplest blocks (tracking, odometry, calibration...) to the full reconstruction formulation, deal with the estimation of poses. Yet the non-convex nature of optimization in the domain of poses in $\mathrm{SE}(3)$ makes that, after decades of research and advances, these problems still remain elusive to general and scalable algorithms which may guarantee the goodness of the obtained solution in the optimal sense.

In this thesis, we address a set of fundamental problems whose core difficulty boils down to precisely optimizing over poses. This includes many geometric 3D registration problems, covering well-known problems with a long research history such as the Perspective-n-Point (PnP) problem and generalizations, extrinsic sensor calibration, or even the gold standard for Structure from Motion (SfM) pipelines: The relative pose problem from corresponding features. Likewise, this is also the case for a close relative of SLAM, Pose Graph Optimization (also commonly known as Motion Averaging in SfM).

The crux of this thesis contribution revolves around the successful characterization and development of *empirically tight* (convex) semidefinite relaxations for many of the aforementioned core problems of 3D Computer Vision. Building upon these empirically tight relaxations, we are able to find and certify the globally optimal solution to these problems with algorithms whose performance ranges as of today from efficient, scalable approaches comparable to fast second-order local search techniques to polynomial time (worst case). So, to conclude, our research reveals that an important subset of core problems that has been historically regarded as hard and thus dealt with mostly in empirical ways, are indeed tractable with optimality guarantees. We truly hope that these successes will motivate further research in this direction, which may radically boost the reliability and performance of current 3D perception algorithms, unleashing the full potential these may have for all kinds of present and future applications.

Acknowledgments

This has been an incredible and long, shared academic journey.

I would like to express my deepest gratitude to my supervisor Javier Gonzalez, for initiating me in the ways of Research, providing me with the opportunities, resources and all necessary support to complete this thesis. As any of my (already PhD) colleagues before me, I think we all owe to Javier his genuine devotion for Research and his ability to transmit this passion for Research to all of us. His interest in exploring new ideas and his unconditional support at all times during this thesis have been, at the very least, noteworthy. I do not think there was a single meeting with Javier where we would not go twice over the expected time to discuss the ongoing topics of research, for there has always been so many exciting items and ideas to cover. Javier, maybe we did not complete the initial goals for what was supposed to be my PhD topic, but I think we both can be definitely proud of what we achieved here.

To everyone else in the MAPIR group, I must say they have been both great colleagues and masters. I want to remark specially the value of doing the thesis surrounded by such a motivated team. I am really glad about the perfect timing, which brought a few of us to work in the exciting field of 3D Computer Vision, bringing a healthy competition spirit which pushed us to do our best and ended up manifesting itself in various successes, with great shared experiences at conferences with great colleagues (most often Ruben Gomez-Ojeda and Mariano Jaimez). I really appreciate the fact that this relationship which got born among MAPIReans during the PhD has outlived this (finite) time and continues as we moved forward to our next stage. With all the general awesomeness in the team, I cannot end this without a few special mentions: To Manu, to whom I thank (and at the same time blame) for introducing me into the infinite world of Linux, opening a whole new landscape that changed my perspective of computing, programming, and the way I work. To Javi, for all that shared commuting, and for enduring hours and hours of the heaviest and most abstract Math talk. And to the MAPIR sports section, for proving me wrong: I did enjoy skiing after all.

A last thank to Jose-Luis Blanco (aka JL), who we consider a perpetual member of the MAPIR team through his contributions, and whose inspiration endures and also lives in this thesis. Even if I did not have the opportunity to directly work with Jose Luis as he had already left the University of Malaga by the time I joined the MAPIR group, I think there is no discussion as for the everlasting presence of JL's work and impact in the team, partly through the MRPT library that still lives at the core of MAPIR, partly because of the many myths around JL unprecedented programming skills, SLAM genius and ability to bear with little sleeping. The opportunities to chat we had in the last years and the praise I got from him regarding some of the contributions in this thesis were, without a doubt, a reinforcing motivation along the long way of the PhD. So, to JL I owe that additional point of motivation to keep exploring the edge of the unknown in SLAM and Computer Vision.

As for any thesis, an important part of my research and my contributions have been strongly supported by the previous work of great researchers. I would like to explicitly mention Nicolas Boumal and Bamdev Mishra, whose library Manopt has been instrumental for exploring and pursuing many of the groundbreaking results of this thesis. Most importantly, I would like to remark their willingness to assist and help with all optimization questions (beyond the use of Manopt) that I might have during my first steps in this exciting topic. I hope this does not trigger too much spam on your mailbox from more junior researchers, but I think you really deserved this explicit mention to your awesome willingness guys!

And now switching to Spanish for my closest ones:

Imposible sería completar esta tesis sin agradecer todo a mi familia primero. A mi familia original, y a mi familia extendida. Por su infinito apoyo y su infinita paciencia. Y porque la alegría por cada pequeño logro que me ha traído hasta aquí tan solo se multiplica cuando puedes compartirla con aquellos que te quieren y a los que quieres.

A mi padre, porque para mí el camino que lleva a la conclusión de esta tesis no comenzó en mi cabeza, sino en la suya: una mente única, curiosa, y extraordinaria. Él me ensenó a través del ejemplo (y de compartir) el placer de la curiosidad, de descubrir y de aprender. Él supo abrazar la curiosidad infinita natural de un niño que admiraba a su padre y que gustoso se pasaría las horas de la noche contemplando las estrellas e intentando desentrañar los misterios más profundos y lejanos del universo juntos, y *cultivó* en mí la pasión por aprender, por comprender e ir más allá.

A mi madre por su amor, cariño y sana preocupación incondicional de madre, que no cambia ni espero que cambie nunca. Pero más allá de eso, quiero remarcar que además de ser madre, siempre has sido una persona fuerte, muy fuerte, capaz de superarse y perserverar ante todo. Si algo me ha enseñado la vida y la perspectiva que sólo los años te dan, es que nos parecemos mucho más de lo que pensaba, y no puedo estar más orgulloso de ello. Gracias por tu ejemplo.

A mi hermana, quien quizás por la diferencia de edad, o por sus muchísimas virtudes, ha sido siempre un referente y ha jugado un papel fundamental para mí. Considero que muchísimo de lo que tengo y he conseguido, en todos los aspectos, te lo debo a tí. Y porque esa deuda es infinita, no se me ocurre mejor forma de pagarla que compartiendo contigo cada uno de estos buenos momentos y los que quedan.

A mi abuelo Francisco, quien a pesar de no haber podido compartir suficientes años conmigo, creó la huella por la que la vida de mi padre, y ahora la mía, discurren. A mi abuela María, sin la cual la familia no estaría ni la mitad de cerca y unida, y quien celebra cada uno de nuestros logros con gran orgullo. A mi abuelo Juan, quien por casualidades de la vida compartió mi mismo destino actual. Me habría encantado poder compartir esto y más historias con él. Y a mi abuela Fuensanta, por demostrarme que nunca es tarde para abrazar la chispa de la vida.

Y a todos los demás, tíos y primos, más o menos cercanos en la distancia, pero siempre cerca en el corazón, de nuevo no sería quien soy ni estaría aquí sin vosotros.

Por último, y no por ser menos importante sino al contrario por merecer su propia sección, si no toda una tesis, quiero agradecer *todo* a mi maravillosa esposa Estefanía. Ella me ha acompañado en todo este camino, *end-to-end*, soportando los momentos y etapas más duros como deadlines, divagaciones, estar ausente, listas infinitas de TODOs... En definitiva todo aquello que típicamente acompaña a un doctorando. Pero por suerte también me ha acompañado en todo lo bueno de ello: celebrando cada logro y publicación, compartiendo cada viaje a conferencias, descubriendo rincones exóticos del planeta, ... Y si bien he de remarcar todo lo que completar una tesis me enseñó (ante todo, me enseñó que apenas sabía nada), he de decir que todo lo demás que aprendí estos años lo aprendí contigo. Y no se me ocurre mejor persona para compartir el camino de la vida. A estas alturas, dudo que nadie en mi vida me haya hecho reír y me haya dado tantos buenos momentos como tú. Gracias por tu sonrisa incondicional. Gracias por tu energía y tus ganas de vivir. Gracias, porque la vida no sería la mitad de lo que es sin tí.

Y a los que vengan o vendrán, sólo decirles: La curiosidad convierte cualquier rincón en un espacio infinito de oportunidades.

Jesús Briales García
Zurich
April 2021

Resumen

1 Introducción

Hemos visto cambios en el mundo, a través de la historia y de nuestros propios ojos. El mundo está cambiando. Pero el mundo, oh amigo mío, el mundo está a punto de cambiar mucho más.

Tecnologías actuales y futuras

Piensa en toda la tecnología involucrada en nuestra vida cotidiana.

Más de cien años de historia de producción industrial en masa respaldan ese automóvil que conduces a casa.

El siglo pasado también vio el surgimiento de los computadores u ordenadores, sólo superado por el desarrollo de la computación móvil, y en concreto el teléfono inteligente ("smartphone"). Puedo pensar fácilmente en varias personas que no tienen automóvil. Pero, ¿puedes pensar en alguien que no tenga su propio teléfono inteligente?

Lo que físicamente podría parecer solo una superficie 2d que ilumina un montón de píxeles se ha convertido para el ciudadano medio en una ventana que le permite acceder a un mundo virtual, Internet, donde las capacidades de comunicación e información son prácticamente ilimitadas.

Si el impacto del teléfono inteligente en nuestra vida diaria es evidente, todavía estamos avanzando activamente hacia tecnologías aún más impactantes. Considera cada vez que has usado tu teléfono inteligente hoy, todas las acciones que has realizado con él y toda la información transferida en el proceso: Comprobar el tiempo, usar su aplicación de mapas favorita para llegar a un destino, enviar mensajes a sus amigos, su familia, revisar el correo ... Ahora imagina obtener toda esa información en un abrir y cerrar de ojos. Pero podemos llegar aún más lejos: ¿Quieres ver en la oscuridad? ¿Quieres compartir la increíble vista del Gran Cañón con tus padres, en tiempo real, como si estuvieran allí? ¿Necesitas gafas? ¿Quieres algunas notas históricas sobre ese monumento que estás mirando? Hablando con alguien en un pub concurrido, ¿te gustaría cancelar el ruido ambiente? Te presento la Realidad Aumentada (ver Fig. 1).

Un éxito más del siglo pasado, los robots han demostrado ser muy valiosos en el entorno industrial para aliviar el trabajo humano mediante la automatización de tareas pesadas, tal vez incluso realizando tareas que están fuera del alcance de las personas. También hemos tenido

Figure 1: Algunos ejemplos del potencial en la Realidad Aumentada. Arriba: la realidad tal como la percibes hoy. Abajo: la realidad percibida a través de gafas de Realidad Aumentada. Imágenes tomadas de la URL https://www.oculus.com/blog/inventing-the-future/.

la oportunidad a estas alturas de ver algunos de estos robots desplegados en casa. Un ejemplo destacado son las aspiradoras robotizadas, con el objetivo de liberarnos de algunas cargas domésticas desde 2002 (Fig. 2).

Más inteligente

Ahora, imagina todo ello hecho *más inteligente*:

Ha sido un día largo, tal vez incluso has decidido compartir una merecida cerveza con tus compañeros después de la oficina. ¿Qué tal si te sientas en tu coche, te relajas y dejas que te lleve a casa (en lugar de al revés)?

Pero las tecnologías inteligentes tienen el potencial para tener aún más impacto en nuestra existencia. Si la evolución ha hecho maravillas a lo largo de millones de años trayendo todas esas capacidades "mágicas" que cada persona ya tiene para sentir y comunicarse, la realidad aumentada trae la promesa de enriquecer nuestras capacidades aún más y, especialmente, nuestra experiencia al interactuar con el mundo virtual, al cual accedemos hoy en día principalmente con nuestros smartphones. ¿Te gustaría saber que viene a continuación? Imagina que todas estas capacidades "añadidas" a nuestros sentidos que describimos anteriormente aparecen automáticamente, según tu contexto, como una extensión natural de tus capacidades y tu voluntad. Te presentamos la Realidad Aumentada *Inteligente*, te presentamos al "superhombre". La Realidad Aumentada Inteligente no consiste solo en poder mostrar elementos virtuales para tus ojos que mejoren tus sentidos o capacidades, también se trata de percibir y comprender el mundo tal como lo haces tú para tomar decisiones fundamentadas y ser proactiva en cómo servirte.

Figure 2: iRobot introdujo la serie roomba en 2002. Imagen tomada de `https://www.irobot.ie/`.

Y ¿qué hay sobre el bastión final de una vida de ciencia ficción? Si la realidad aumentada tiene el poder de mejorar nuestras capacidades cognitivas, proveyendo *información*, en la realidad física la robótica tiene la capacidad de *actuar* por nosotros, automatizando muchas labores manuales. Pero podemos ir mucho más allá (Fig. 3): robots para zonas de desastre, drones... Todos tienen el potencial de realizar importantes tareas, para nosotros. Pero para que un robot sea tan capaz como un humano, ¿no debería tener capacidades similares?

¿Qué se necesita?

A estas alturas del discurso, espero que el futuro te parezca emocionante y prometedor, a mí me lo parece. Existen muchos desafíos en numerosas disciplinas y campos que se interponen en el camino para alcanzar el futuro que esbozamos arriba. Pero si hay algo sorprendente en todo esto es que hay muchas razones y hechos para pensar que se puede conseguir. Sin embaro, es sólo la perseverancia, la inversión de una importante cantidad de recursos y, sobre todo, el genio humano, pueden hacer realidad estos escenarios.

Pero, en un sentido más técnico, ¿qué necesitamos para llegar allí? Como se mencionó anteriormente, todas y cada una de estas aplicaciones son multidisciplinares y requieren numerosas herramientas. Pero aquí queremos llamar nuestra atención sobre una necesidad fundamental que comparten todas las tecnologías mencionadas: la necesidad de una Percepción Artificial Inteligente.

El campo de la Inteligencia Artificial está en auge. Pero para que una máquina nos ayude en nuestra vida diaria (en cualquiera de las formas descritas anteriormente), no solo necesita ser inteligente, sino también *percibir* nuestra vida diaria de la misma manera que lo hacemos

Figure 3: Izquierda: un (hipotético, futuro) robot humanoide (imagen tomada de `https://blogs.3ds.com/northamerica/future-robots-and-ensuring-human-safety/`). Derecha: un dron operando en un entorno peligroso (imagen tomada de `https://newsroom.cisco.com/feature-content?type=webcontent&articleId=1940944`).

nosotros. Tan sólo así, un agente automatizado podrá tener las herramientas para comprender el entorno lo suficientemente bien como para saber qué se debe hacer y cuándo, de manera proactiva. Sólo de esa manera, cualquiera de las capacidades anteriores puede convertirse en parte integral de nuestra vida.

Si algo define las capacidades de un ser humano para interactuar y comprender el entorno, es (entre otras cosas, por supuesto) su capacidad para percibir la estructura tridimensional del mundo en el que vivimos.

Por suerte, la madurez alcanzada por muchos de los (ahora clásicos) algoritmos e ideas en áreas como la Geometría desde Múltiples Vistas ("Multiple View Geometry") u Optimización proporcionan a día de hoy una serie de herramientas para percepción 3D sobre las cuales ya se construyen numerosas aplicaciones en la actualidad, y que son la semilla para las capacidades necesarias en muchas de las tecnologías emergentes mencionadas anteriormente.

Analicemos con mayor detalle estas herramientas fundamentales de percepción 3D.

2 Una herramienta fundamental de la Visión Artificial en 3D (3D Computer Vision): el "Ajuste de Paquetes" ("Bundle Adjustment")

De las muchas técnicas que permiten la percepción 3D, el estándar es probablemente la "optimización por lotes" ("Batch Optimization"), también conocida como "ajuste de paquetes" ("Bundle Adjustment", BA), "estructura a partir del movimiento" ("Structure from Motion", SfM) o "localización y mapeo simultáneos" ("Simultaneous Location and Mapping", SLAM). Existen algunas distinciones clásicas entre estos métodos, pero en la práctica estos límites se han vuelto menos claros con los avances en los distintos campos. En términos muy generales, esta técnica intenta encontrar la geometría 3D del mundo circundante, utilizando la información proporcionada por un conjunto de sensores (o dispositivo), que puede variar desde una sola cámara monocular hasta sistemas multicámara más complejos, a menudo acompañados de sensores de rango o Unidades de Medición Inercial ("IMUs") (ver Fig. 4). Intrínsecamente, el

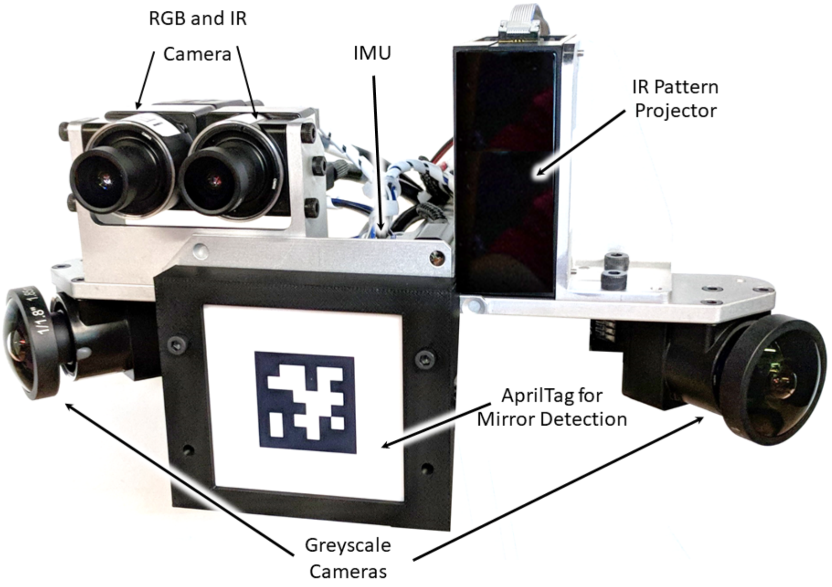

Figure 4: Equipo de captura profesional empleado en [2], compuesto por cámaras en escala de grises, RGB e infrarrojas, junto con Unidades de Medición Inercial (IMU) y un proyector infrarrojo.

problema también implica estimar el movimiento de los sensores y, a menudo, refinar muchos de los parámetros internos de estos sensores.

Este tipo de técnicas, con los datos adecuados, pueden conducir a resultados sobresalientes (ver Fig. 5).

En un tono más técnico, los métodos "estado del arte" plantean este problema como una instancia de Estimación de Máxima Verosimilitud ("Maximum Likelihood Estimation", MLE), donde asumiendo una cierta distribución de probabilidad para el ruido de los sensores buscamos un modelo que maximice la probabilidad general de toda la información disponible. Esta es una formulación muy atractiva de cualquier problema probabilístico desde un punto de vista teórico, dadas las propiedades y garantías que conlleva la estimación de máxima verosimilitud (MLE). Más allá de la "elegancia" de este enfoque, esto eventualmente se reduce a resolver un problema de optimización en el que buscamos los parámetros del modelo que maximizan la consistencia de todos los datos del sensor disponibles, o de manera equivalente minimizar el error de las predicciones del modelo respecto a los valores reales observados en nuestro sensores. Sin embargo, esto en realidad implica una tarea computacional fundamentalmente difícil. Para entender por qué, primero debemos hacer una revisión rápida de la optimización y sus conceptos básicos.

2.1 Nociones básicas de optimización

Al resolver un problema mediante optimización, nuestro objetivo es encontrar el valor más bajo posible para una función en una determinada región o dominio del problema.

Los métodos de optimización que se encuentran típicamente en la visión geométrica por computador y en "Bundle Adjustment" funcionan de manera iterativa, realizando una búsqueda local desde un punto inicial con el objetivo de llegar finalmente a un punto crítico (mínimo

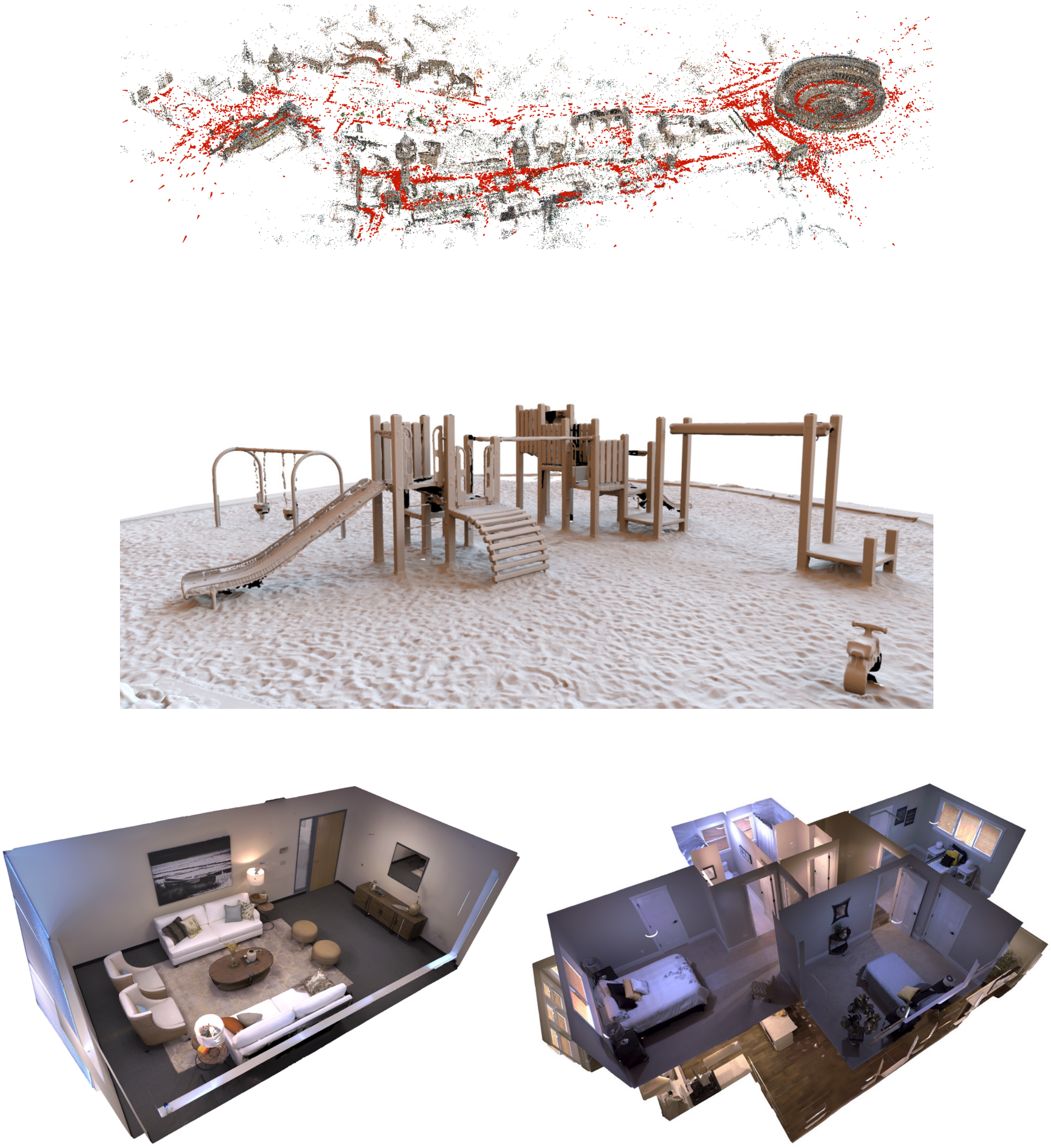

Figure 5: Las técnicas de visión geométrica por computador de última generación junto con los datos proporcionados por diferentes sensores pueden proporcionar reconstrucciones 3D de muy alta calidad. Arriba: reconstrucción dispersa ("sparse") mediante SfM basada en imágenes de turismo fotográfico [3]. Enmedio: reconstrucción de alta fidelidad con tecnología LIDAR [4]. Abajo: reconstrucciones de alta fidelidad realizadas con equipo profesional en Fig. 4 [2].

local) en la función a optimizar. Si el punto crítico encontrado es el óptimo globalmente, lo logramos. Estos métodos son notablemente rápidos en la mayoría de los casos de interés, lo cual es una virtud esencial para la mayoría de las aplicaciones donde la Visión Artificial en 3D es un componente central.

Si hay *un solo* punto crítico en todo el dominio, el problema será fundamentalmente fácil, ya que solo necesitamos seguir iterando en nuestro método hasta que finalmente lleguemos a este punto crítico. Esto es básicamente lo que define un *problema de optimización convexa*: tener un mínimo local (y por tanto global) único.

Pero ¿y si existen múltiples puntos críticos? En este escenario, un método iterativo típico podría estancarse en los mínimos locales incorrectos, impidiéndonos encontrar la solución realmente buscada que se encuentra en el mínimo global. Esto es lo que define un *problema de optimización no convexo*.

Así, en este escenario, las alternativas más comunes son mantener el mejor mínimo local encontrado por un método iterativo local (sin más garantías sobre su optimalidad global), o recurrir a métodos exploratorios más complicados (si buscamos garantías de optimalidad global), como por ejemplo "Branch-and-Bound", que básicamente hace algunos trucos inteligentes para particionar de forma recursiva el dominio y verificar si una partición puede contener el óptimo global. Esta última familia de métodos, sin embargo, es intrínsecamente mucho más costosa computacionalmente que los métodos iterativos locales rápidos y escalables.

2.2 De vuelta al metodo de "Bundle Adjustment" como problema de optimización

Ahora estamos en condiciones de entender por qué resolver el problema de "Bundle Adjustment" es un problema tan difícil. El problema de "Bundle Adjustment" es, por definición, un problema de optimización no lineal, no convexo y de dimensiones enormes. Debido a esto, ningún algoritmo existente puede resolver una instancia del problema de "Bundle Adjustment" común con garantías de optimalidad global.

Así, en la práctica, en lugar de encontrar la estimación de máxima verosimilitud (MLE), los métodos prácticos se enfocan en encontrar un *punto crítico* del problema de optimización subyacente, con la *esperanza* de que este punto crítico será de hecho la solución globalmente óptima, y por lo tanto corresponderá a la estimación de máxima verosimilitud deseable. Esta tarea ha sido abordada por muchas de las soluciones actualmente existentes mediante la aplicación de métodos rápidos de optimización numérica de primer o segundo orden, que es un enfoque atractivo por las muchas ventajas que proporciona junto con técnicas de optimización de última generación, como velocidad de convergencia rápida en la proximidad de un punto crítico o escalabilidad en el tamaño del problema aprovechando que el problema es "disperso" ("sparse"). En general, esto ha resultado en un conjunto de herramientas y librerías [5,6] que pueden abordar el problema de "Bundle Adjustment" de una manera eficiente y escalable, y que funcionan bien bajo el supuesto de comenzar *cerca de la solución óptima*, debido a la naturaleza iterativa de estas técnicas.

3 Límites de usabilidad para la Visión Artificial en 3D estado del arte

En medio del fuerte aumento en las aplicaciones relacionadas con la Inteligencia Artificial (IA), a medida que aumentamos las capacidades de la IA o la automatización habilitada por la percepción y, por lo tanto, la cantidad de tareas que podemos asignarles, también ha habido una preocupación motivada por la resiliencia y fiabilidad de estos sistemas. Para muchas aplicaciones con reciente éxito (aplicaciones móviles, etc.), los errores pueden provocar una experiencia de usuario no tan fluida, que finalmente termina con una aplicación no completamente estable y un cliente no tan satisfecho. Pero para una gran cantidad de aplicaciones de interés que propulsan este campo, como la conducción autónoma o la robótica, un error del sistema (debido, por ejemplo, a fallos en un algoritmo) puede incurrir en altos costes que van desde daños a la propiedad hasta la pérdida de vidas humanas [7]. Por lo tanto, para alcanzar todo el potencial de estas técnicas, debemos posibilitar la adopción total en estos escenarios críticos proporcionando un grado adicional de solidez, fiabilidad y comprensión de las técnicas que se utilizan.

En el contexto de la geometría 3D (y por lo tanto en las herramientas de reconstrucción 3D actuales, como el problema de "Bundle Adjustment"), un punto crítico subóptimo puede estar arbitrariamente lejos de la solución correcta, por lo que encontrar la solución *realmente óptima* a la formulación MLE se convierte en un *requisito*. Dado que, como vimos, las técnicas de "Bundle Adjustment" proceden principalmente a través de métodos de optimización iterativos, obtener una estimación inicial suficientemente buena para comenzar el refinamiento iterativo subsiguiente es un paso crucial hacia el éxito de toda la reconstrucción 3D. Para este propósito, la mayoría de los sistemas constan de varios pasos en los que se aprovechan las soluciones a los subproblemas más simples para construir de forma incremental una estimación inicial que pueda incorporarse al esquema de refinamiento final.

Se han propuesto numerosas soluciones para estos diversos subproblemas clásicos fundamentales en el contexto del problema de "Bundle Adjustment". Desafortunadamente, incluso estas versiones simplificadas del problema de "Bundle Adjustment" todavía están lejos de ser triviales. Así, muchos de estos algoritmos de inicialización básicos que tienen como objetivo asegurar el éxito del ajuste posterior del paquete siguen siendo de naturaleza principalmente empírica. Esto hace que toda el sistema sea más frágil, en el sentido de que incluso si muchas de las soluciones actuales funcionan perfectamente en circunstancias favorables, pueden fallar catastróficamente en un escenario ligeramente menos favorable. En última instancia, este aspecto empírico de muchas soluciones existentes puede hacer que el régimen de operación fiable de los algoritmos sea más limitado de lo que nos gustaría. Peor aún, puede haber poca o ninguna certeza sobre cuál es el régimen seguro de operación.

Solo lentamente se están realizando más investigaciones sobre estos bloques fundamentales de la Visión Artificial en 3D, arrojando algo de luz sobre cuáles son los límites fundamentales y qué garantías y fiabilidad podemos obtener de los métodos existentes.

Así que, en conclusión, la fiabilidad es una característica crítica que restringe un uso más amplio de las herramientas actuales de percepción 3D. Si bien la fiabilidad puede (y debe) abordarse de varias maneras (por ejemplo, mediante una mayor redundancia de sensores, real-

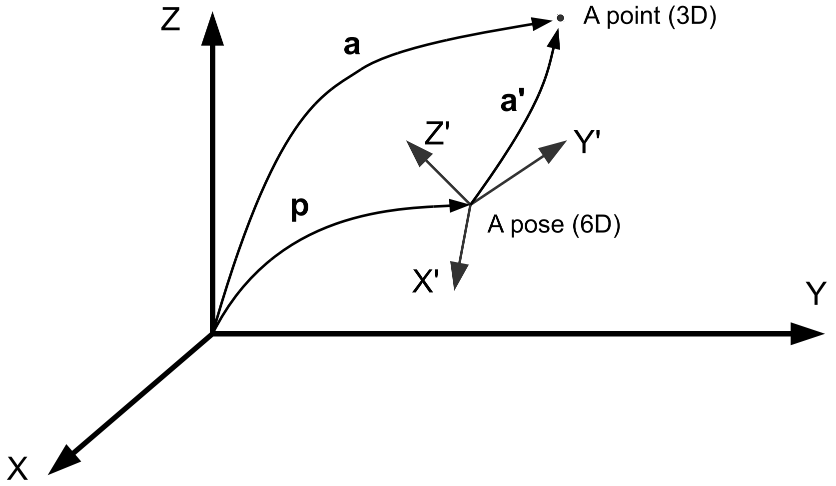

Figure 6: Una pose como se presenta en [8]

izando verificaciones adicionales, etc.), se puede ganar mucho mejorando las garantías de los métodos actuales. Así, en esta tesis, nuestra principal contribución se enmarca en la mejora de la comprensión y las garantías a la hora de resolver varios subproblemas fundamentales en la visión por ordenador en 3D.

4 Algunos subproblemas fundamentales en la Visión Artificial en 3D

A continuación, presentaremos dos tipos de problemas que juegan un papel fundamental para inicializar el problema de "Bundle Adjustment". En esencia, ambos comparten un objetivo común (y, como veremos más adelante, un desafío): la estimación de movimiento o poses. Pero primero, ¿qué es una pose?

¿Que es una pose?

Tiene sentido que para comprender la forma tridimensional del mundo, un ingrediente esencial sea describir dónde están ubicadas las cosas. Para un objeto rígido esto involucra dos parámetros: su posición y su orientación (o equivalentemente su rotación). La combinación de posición y orientación se denomina *pose* de un objeto. Más allá del concepto de alto nivel, en la práctica, esta descripción debe darse respecto a algún sistema de coordenadas de referencia (de aquí en adelante, el sistema de coordenadas global).

Dado un sistema de coordenadas global y el sistema de coordenadas local de un objeto (Fig. 6), la pose del objeto respecto al sistema global básicamente describe el sistema de coordenadas local del objeto en términos de coordenadas globales.

El concepto de pose también subyace a la idea de movimiento, y así se remonta a los conceptos fundamentales de la mecánica clásica, conectándose con la noción de transformación de un cuerpo rígido. Por lo tanto, el trabajo y cálculo con poses también es fundamental para muchas otras disciplinas, como la robótica.

Problemas geométricos de registro de pose por parejas en 3D ("Geometric 3D Pairwise Pose-Registration problems")

En los problemas de registro, como en el problema de "Bundle Adjustment", nuestro objetivo es unificar un conjunto de observaciones de un modelo común de manera coherente. Sin embargo, en este caso, es común introducir varias simplificaciones que reducen parte de la enorme complejidad en el problema de "Bundle Adjustment" completo más general. Así, mientras que en el problema de "Bundle Adjustment" es normal trabajar en espacios complejos y significativos, como puntos 2D en la imagen proyectados desde la cámara, o incluso valores fotométricos en la imagen, en los problemas de registro se acostumbra trabajar con características geométricas en 3D y elementos geométricos primitivos (puntos, líneas, planos). También, mientras que en el problema de "Bundle Adjustment" a menudo optimizamos una gran cantidad de parámetros que incluyen el modelo 3D del entorno, los parámetros internos del sensor, etc., en los problemas de registro nos limitamos a estimar la mejor pose para que dos conjuntos de características geométricas coincidan[1], asumiendo que el resto de los parámetros son lo suficientemente buenos como para que realmente pueda ocurrir un buen ajuste.

A pesar de las simplificaciones anteriores, este tipo de problemas de registro son ubicuos y subyacen a muchos subproblemas centrales en la visión clásica por computador en 3D, desempeñando un papel fundamental para iniciar algoritmos más complejos, como el problema de "Bundle Adjustment". Algunos ejemplos incluyen:

- Calcular la pose absoluta de un sensor visual o de rango con respecto a un modelo geométrico dado (seguimiento o "tracking").
- Calcular el movimiento relativo de uno de estos sensores entre dos momentos (odometría).
- Calcular la calibración extrínseca (pose relativa) de dos de estos sensores en un equipo (calibración).

Las características geométricas 3D son omnipresentes cuando se trabaja con el tipo de sensores anteriores. Por ejemplo, los puntos y líneas 2d en las imágenes de la cámara se retroproyectan como líneas y planos 3D cuando se considera la geometría proyectiva de la cámara; los sensores de rango generalmente detectan puntos 3D en los que podemos ajustar planos 3D, que pueden cruzarse a su vez formando líneas 3D. Por tanto, cualquiera de los problemas anteriores se puede plantear cuidadosamente como el registro de dos conjuntos de primitivas geométricas 3D (puntos, líneas, planos).

Optimización de Grafo de Poses ("Pose Graph Optimization")

Como consecuencia de la omnipresencia de los problemas anteriores, muy a menudo nos encontramos con información redundante sobre el movimiento relativo o la pose de varios objetos a lo largo del tiempo (a pares). El uso de toda esta información redundante disponible para estimar un modelo global único conlleva un problema matemático que involucra solo poses, que ha sido

---

[1] En realidad, un problema de registro puede consistir en más de dos conjuntos de características para encajar, pero por simplicidad restringiremos nuestro discurso aquí al caso base de dos conjuntos.

durante mucho tiempo objeto de estudio en diversas comunidades: "Optimización de Grafo de Poses" ("Pose Graph Optimization"), "Promediado del Movimiento" ("Motion Averaging"), "Sincronización de Poses" ("Pose-Synchronization")... se refieren todos al mismo concepto y problema subyacente. Tenga en cuenta que mientras que los problemas de registro presentados anteriormente limitaban significativamente la complejidad del problema al considerar una sola pose para optimizar, en el problema de "Pose Graph Optimization" podemos tener hasta decenas de miles de poses para optimizar (o incluso más). Por lo tanto, incluso si el problema de "Pose Graph Optimization" sigue siendo una fuerte simplificación del problema de "Bundle Adjustment" o SLAM completo más general, ya se encuentra significativamente más cerca en términos de escalabilidad y complejidad del dominio.

Un desafío subyacente común: optimizar poses

A pesar de eliminar muchas capas de complejidad con respecto al problema de "Bundle Adjustment" completo, incluso en el más simple de los escenarios anteriores, sigue existiendo un desafío fundamental: Optimizar sobre una pose.

Desde una perspectiva matemática, las poses (o equivalentemente "transformaciones rígidas") son elementos del *grupo Euclideo especial* SE(3), uno de los grupos más antiguos y estudiados en matemáticas (ya mucho antes de que se inventara el concepto de grupo). Por lo tanto, podría sorprender que lidiar con poses pueda ser aún un desafío en la actualidad, ¡pero lo es!

La razón de esto se remonta a los conceptos básicos de optimización y las dificultades que implica la no convexidad. El espacio de rotaciones 3D (recuerde que una pose 3D consiste en una posición y una orientación) es *no convexo*, y así la optimización de cualquier función objetivo (por simple que sea) puede presentar múltiples mínimos locales. Por lo tanto, cualquier problema de optimización general que implique una pose no es convexo y, por lo tanto, no es trivial de resolver.

Como consecuencia, cualquiera de los problemas considerados anteriormente es en general difícil de resolver con garantías globales de optimalidad, aunque siempre existen excepciones. Por ejemplo, existe una solución de forma cerrada basada en una descomposición de valores propios para el problema de Procrustes entre dos nubes de puntos [9, 10]. También para algunos de estos problemas, podemos encontrar todos los mínimos locales utilizando métodos estado del arte, pero esta estrategia no escala bien a medida que aumenta la complejidad del problema y el número de mínimos locales puede explotar.

Hasta el momento de las contribuciones en esta tesis, las pocas soluciones existentes que se ocupan de la optimización sobre poses de una manera certificable estaban muy lejos del ámbito de la computación en tiempo real, proporcionando optimización principalmente a través de técnicas exploratorias (como "Branch-and-Bound") [11, 12].

5 Contribuciones y esquema de esta tesis

En el contexto de los bloques fundamentales de la Visión Artificial 3D antes mencionados, esta tesis tiene como objetivo caracterizar diferentes subproblemas cuya dificultad central proviene de la estimación de poses (Sec. 3). Esta tesis contribuye a superar las limitaciones de fiabilidad actuales para obtener soluciones óptimas garantizadas dada la naturaleza no convexa de esos problemas, mediante la exploración de enfoques prácticos para obtener una solución óptima a los subproblemas geométricos relevantes.

Los aportes de esta tesis se pueden dividir en varios bloques.

5.1 Abordar diferentes tareas de Visión Artificial en 3D como problemas de registro

En la primera parte de esta tesis, exploramos el potencial de los problemas de registro geométrico para modelar y abordar algunas tareas típicas de la Visión Artificial en 3D (Sec. 3).

Una de esas tareas es la calibración extrínseca de una cámara monocular y un sensor de rango láser 2D (o LIDAR), asumiendo una calibración intrínseca ya conocida para ambos. Debido a la escasa superposición en el tipo de datos percibidos por estas dos modalidades de sensor, construir la asociación de datos necesaria para realizar la calibración extrínseca no es sencillo. En nuestro artículo en ICRA15 [13] propusimos un enfoque novedoso que supera esta limitación aprovechando las limitaciones estructurales presentes en cualquier entorno tipo "Manhattan" típico de las estructuras creadas por el hombre para establecer correspondencias entre la retroproyección en 3D de las características extraídas en la imagen de la cámara (planos y líneas) y las extraídas en el escaneo LIDAR (líneas y puntos), evitando la necesidad de construir un patrón de calibración específico para la tarea. En nuestro artículo en IROS15 [14] aportamos una solución de forma cerrada para el caso mínimo del problema de registro anterior, que básicamente se reduce a una observación completa de un trihedro ortogonal tanto desde la cámara monocular como desde el sensor de rango láser 2D.

La segunda tarea abordada es la de estimar la orientación de una cámara monocular a partir de la observación de líneas rectas en (nuevamente) un entorno "Manhattan" creado por el hombre. En nuestro artículo de revista en JMIV [15], proporcionamos una solución simple y rápida al problema de registrar tres planos retroproyectados en el sistema de coordenadas de la cámara a las correspondientes líneas 3D en el mundo cuando estas tienen direcciones ortogonales. Entonces usamos esta solución para implementar un giroscopio visual (es decir, estimar las rotaciones relativas de la cámara a partir de las imágenes).

A través de los problemas anteriores, confirmamos las principales afirmaciones en la introducción de esta tesis (Sec. 3): Los problemas que involucran la estimación de pose son muy comunes, sin embargo, excepto por algunos problemas específicos donde se pueden implementar soluciones de forma cerrada, las soluciones óptimas garantizadas para la mayoría de los problemas más generales e interesantes siguen siendo esquivas. Este punto, sin embargo, lo abordamos en el siguiente conjunto de contribuciones.

5.2 Optimalidad global en problemas de registro geométrico 3D

En el segundo bloque de esta tesis, revisamos el tema fundamental del registro geométrico 3D y mostramos que algunos problemas clásicos centrales de la visión 3D por computador que durante mucho tiempo han sido considerados como globalmente intratables (en tiempo polinomial), son de hecho resolubles con garantías de optimalidad mediante una adecuada relajación convexa.

En CVPR17 [16], mostramos que para un problema de registro 3D multimodal bastante general, que involucra correspondencias de puntos a puntos, líneas y/o planos, es posible obtener una solución óptima certificable a través de una relajación convexa apropiada que resulta ser ajustada en virtualmente todos los casos probados.

En CVPR18 [17], abordamos uno de los problemas más clásicos y fundamentales en la geometría de múltiples vistas, el problema del cálculo de la pose relativa a partir de correspondencias 2D (para cámaras calibradas). Nuevamente, en este caso podemos caracterizar una relajación convexa del problema que resulta ser ajustada o "tight" en prácticamente todos los escenarios probados, proporcionando un procedimiento para obtener una solución óptima certificable en este problema central de la Visión Artificial 3D.

5.3 Optimalidad global en la optimización de grafo de poses

Siguiendo con el tema de la estimación de pose y la dificultad inherente de los problemas relacionados, en la última parte de esta tesis vamos más allá de los problemas de registro por pares y nos enfocamos en construir un método óptimo y certificable para el problema de "Pose Graph Optimization" (Sec. 3). Para ello, en primer lugar, caracterizamos en nuestro artículo en IROS16 [18] una relajación convexa novedosa a través del Lagrangian dual del problema de "Pose Graph Optimization" que se vuelve empíricamente ajustado en circunstancias comunes. Cuando la relajación es ajustada o "tight", esta permite certificar la optimalidad de una solución candidata al problema de "Pose Graph Optimization". Gracias a un tratamiento más cuidadoso de la estructura del problema, el algoritmo de certificación resultó en un procedimiento mucho más limpio y rápido que para el método estado del arte contemporáneo [19].

Sobre la base de esta misma relajación convexa, en nuestro artículo en ICRA17 [20] contribuimos un método unificado que proporciona directamente la solución globalmente óptima certificable para una instancia del problema de "Pose Graph Optimization" desde la solución a su problema dual, si es ajustada, o una inicialización empírica competitiva con otras alternativas de lo contrario en aquellos escenarios extremos donde el problema dual no es ajustado, por ejemplo cuando los datos se vuelven muy ruidosos.

Por último, en la revista RAL17 [21] combinamos la relajación convexa para el problema de "Pose Graph Optimization", caracterizada en nuestro trabajo anterior, con avances de última generación en métodos de optimización numérica en geometría diferencial ("manifolds") para producir un método de regiones de confianza de segundo orden ("second-order trust-regions method") rápido y certificable que es capaz de converger globalmente a la solución óptima de una instancia del problema de "Pose Graph Optimization" cuando su problema dual Lagrangiano es ajustado.

6 Publicaciones

La presente tesis integra las siguientes publicaciones:

Revistas

- J. Briales and J. Gonzalez-Jimenez, **"A Minimal Closed-form Solution for the Perspective Three Orthogonal Angles (P3oA) Problem: Application To Visual Odometry,"** in Journal of Mathematical Imaging and Vision, vol. 55, no. 3, pp. 266-283, 2015, doi: 10.1007/s10851-015-0620-x. **JCR in Computer Vision and Pattern Recognition: Q1, T1**

- J. Briales and J. Gonzalez-Jimenez, **"Cartan-Sync: Fast and Global SE(d)-Synchronization,"** in IEEE Robotics and Automation Letters (RA-L), vol. 2, no. 4, pp. 2127–2134, 2017, doi: 10.1109/LRA.2017.2718661. Not indexed yet in 2017.

Conferencias Internacionales

- R. Gomez-Ojeda, J. Briales, E. Fernandez-Moral, and J. Gonzalez-Jimenez, **"Extrinsic calibration of a 2d laser-rangefinder and a camera based on scene corners,"** in IEEE International Conference on Robotics and Automation (ICRA), 2015, doi: 10.1109/ICRA.2015.7139700. **Conference Ranking: A1 (Qualis).**

  Tanto el primer como segundo autor tuvieron el mismo grado de contribución en este trabajo.

- J. Briales and J. Gonzalez-Jimenez, **"A minimal solution for the calibration of a 2D laser-rangefinder and a camera based on scene corners,"** in IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2015, doi: 10.1109/IROS.2015.7353625. **Conference Ranking: A1 (Qualis), A (ERA).**

- J. Briales and J. González-Jiménez, **"Fast Global Optimality Verification in 3D SLAM,"** in IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2016, doi: 10.1109/IROS.2016.7759681. **Conference Ranking: A1 (Qualis), A (ERA).**

- J. Briales and J. González-Jiménez, **"Initialization of 3D Pose Graph Optimization using Lagrangian duality,"** in IEEE International Conference on Robotics and Automation (ICRA), 2017, doi: 10.1109/ICRA.2017.7989600. **Conference Ranking: A1 (Qualis).**

- J. Briales and J. González-Jiménez, **"Convex Global 3D Registration with Lagrangian Duality,"** in IEEE Conference on Computer Vision and Pattern Recognition, 2017, doi: 10.1109/CVPR.2017.595. **Conference Ranking: A1 (Qualis), A (ERA).**

- J. Briales, L. Kneip, and J. Gonzalez-Jimenez, **"A certifiably globally optimal solution to the non-minimal relative pose problem,"** in IEEE Conference on Computer Vision and Pattern Recognition, 2018, doi: 10.1109/CVPR.2018.00023. **Conference Ranking: A1 (Qualis), A (ERA).**

7 Contexto y línea temporal

Esta tesis es el resultado de cinco años de trabajo del autor en el grupo de investigación *Machine Perception and Intelligent Robotics (MAPIR)*[2], a partir de octubre de 2013 cuando me incorporé al grupo MAPIR como asistente de investigación, ya bajo la supervisión del Prof. Javier González Jiménez. Después de esto, recibí una beca FPU (*Formación de Personal Universitario*), con este mismo grupo, soportada por el Ministerio de Educación español, que financió principalmente esta investigación doctoral. El grupo MAPIR forma parte del departamento de Ingeniería de Sistemas y Automatización, de la Universidad de Málaga, y tiene una amplia experiencia en robótica móvil, visión artificial y olfato robótico.

Durante este período, el autor completó el programa de doctorado en *Ingeniería Mecatrónica* en el departamento de Ingeniería de Sistemas y Automatización. Al comienzo de este viaje, tuve la oportunidad de participar en la *Escuela Internacional de Verano de Visión Artificial* (International Computer Vision Summer School, ICVSS), celebrada en Sicilia en 2016, que brindó una gran oportunidad para jóvenes investigadores y estudiantes de doctorado para reunirse e interactuar directamente y discutir con líderes mundiales en el campo de la Visión Artificial (y de hecho he tenido la oportunidad en los años siguientes de seguir encontrándome a muchas de estas personas una y otra vez, viendo nuestros progresos alineados con el progreso de la disciplina). Más tarde tuve la oportunidad de participar y presentar nuestro trabajo en múltiples conferencias internacionales relevantes sobre robótica y Visión Artificial, en Seattle (Estados Unidos), Hamburgo (Alemania), Daejeon (Corea del Sur), Singapur (Singapur), Honolulu (Estados Unidos), Vancouver (Canadá) y Salt Lake City (Estados Unidos) y el taller internacional sobre líneas, planos y modelos de "Manhattan" para mapeo 3-D (LPM17) celebrado conjuntamente con la conferencia IROS en Vancouver, 2017.

De marzo de 2017 a junio de 2017, el autor completó una estancia de investigación en el *Computer Vision Group* de la Chalmers University of Technology (Gotemburgo, Suecia). De enero de 2018 a julio de 2018, el autor fue becario de investigación en *Oculus (Facebook)* en Redmond (EE. UU.) bajo la supervisión del Dr. Hauke Strasdat, y se incorporó nuevamente como científico investigador a tiempo completo en noviembre de 2018.

Además, el autor ha sido un revisor activo de manuscritos en prestigiosos congresos y revistas, incluyendo *IEEE Conference on Computer Vision and Pattern Recognition* (CVPR), *International Journal of Robotics Research* (IJRR), *IEEE International Conference on Computer Vision* (ICCV), *International Conference on Robotics and Automation* (ICRA), *Pattern Recognition Letters*, *IEEE Robotics and Automation Letters* (RA-L), *IEEE Transactions on Robotics* (T-RO), *International Conference on Intelligent Robots and Systems* (IROS), *Journal of Mathematical Imaging and Vision* (JMIV) o *Robotics: Science and Systems* (RSS).

Durante el desarrollo de esta tesis, el autor impartió varios cursos como ayudante de cátedra en la Escuela Superior de Ingeniería Informática y en la Escuela Técnica Superior de Ingeniería Industrial de la Universidad de Málaga.

---

[2] http://mapir.isa.uma.es/

8 Conclusiones y líneas futuras

En esta tesis hemos abordado múltiples problemas (de optimización) fundamentales en la Visión Artificial en 3D que, a pesar de la madurez relativa del campo, seguían siendo inherentemente difícil de resolver debido a la naturaleza no convexa de la optimización sobre poses en 3D.

Primero, exploramos y mostramos la ubicuidad de los problemas de registro 3D en múltiples formas (que implican el registro de diferentes primitivas geométricas), y cómo estos sirven como el problema central de estimación para resolver diferentes tareas en la Visión Artificial 3D, como la calibración o la odometría (Sec. II).

Luego, pasamos a la tarea de resolver con garantías de optimización global *y* eficientemente algunos de estos problemas de registro 3d genéricos o ubicuos (Sec. III), así como el problema relacionado o complementario de sincronizar/promediar múltiples poses globales basadas en observaciones relativas entre ellas (Sec. IV), que surge naturalmente en muchas aplicaciones al combinar la salida de múltiples problemas de registro 3D. A pesar de la existencia de diferentes enfoques o técnicas para resolver problemas de optimización con garantías de optimización global, por ejemplo técnicas exploratorias como "Branch-and-Bound" (BnB) o técnicas algebraicas basadas en el cálculo de todos los puntos críticos para la función de coste, la caracterización de una relajación convexa ajustada muestra ser uno de los enfoques más efectivos y escalables. En particular, las contribuciones de esta tesis mostraron que la resolución de los "programas semidefinidos" ("Semidefinite Programs", SDP) que surgen de las relajaciones aplicadas a los distintos problemas estudiados superan a las alternativas basadas en BnB para la optimización global genérica sobre poses, incluso si se utilizan solo métodos estándar para resolver estos problemas SDP.

Además, mostramos cómo al aprovechar hábilmente la existencia de estas relajaciones ajustadas o "tight", podemos incluso implementar algoritmos certificables específicos para el problema [21] que compiten con los métodos iterativos locales estado del arte en términos de rendimiento y, al mismo tiempo, proporcionan las garantías de optimalidad global que sólo son posibles a través de la correspondiente relajación convexa ajustada.

Sobre el arte de caracterizar relajaciones ajustadas

Encontrar una relajación convexa ajustada para un problema dado (la pieza clave para la mayoría de nuestras contribuciones) no es, por supuesto, un trabajo trivial. Cualquier tipo o familia de problemas no convexos, y en particular cualquier Programa Cuadrático con Restricciones Cuadráticas (Quadratically Constrained Quadratic Program, QCQP), sigue siendo fundamentalmente difícil (en jerga informática, "NP-hard") de resolver con garantías de optimalidad global, hasta que encontremos un enfoque específicamente elaborado que demuestre lo contrario (ya sea con garantías teóricas o empíricas).

En el enfoque seguido en esta tesis para encontrar relajaciones convexas ajustadas para algunos de estos problemas, nos apoyamos recurrentemente en el hecho de que diferentes formulaciones equivalentes para un problema de optimización dado pueden conducir a relajaciones significativamente diferentes, que van desde relajaciones sin valor práctico alguno hasta relajaciones *ajustadas* [22]. Si bien este aspecto de los problemas de optimización convierte la elaboración de relajaciones alternativas en una especie de *arte* matemático, también es cierto (y se muestra en esta tesis) que ciertos trucos y métodos funcionan bien sobre una variedad de problemas, lo que nos permite explorar relajaciones potencialmente útiles de formas más sistemáticas. Si tuviéramos que destacar una de estas técnicas de entre las que se aprovechan en esta tesis, sería sin duda el gran potencial que tiene introducir restricciones adicionales equivalentes pero independientes al conjunto de restricciones habitual definiendo el dominio del problema. Esta técnica ha sido fundamental para lograr las relajaciones ajustadas que han resultado todo un éxito para los problemas de registro 3D abordados (Sec. II).

Sobre el poder de los datos

Otra observación relevante de nuestro trabajo a través de esta tesis es cómo más datos pueden facilitar los problemas de optimización (sí, los datos no solo son poderosos para el campo del aprendizaje automático o "Machine Learning"). Más datos, lo que significa un mayor número de correspondencias en los problemas de registro, o una mayor conectividad en el grafo de los problemas de sincronización, pueden hacer que estos problemas de optimización sean inherentemente más simples. Esto se refleja en el hecho de que las relajaciones más simples de esas instancias de problemas se vuelven ajustadas y permiten resolver el problema de manera óptima. Esta es nuestra intuición basándonos en la observación de que, incluso si nuestras relajaciones propuestas para los problemas de registro 3D se destacan por su capacidad de permanecer (empíricamente) ajustadas en todos los casos probados, las relajaciones más simples fueron a menudo suficiente para resolver aquellos casos con una gran cantidad de observaciones o correspondencias. También es la intuición tras el hecho de que para la sincronización de poses, a pesar de la gran dimensionalidad y la complejidad general del dominio del problema, una relajación relativamente sencilla del problema resulta ser ajustada en la práctica, sin la necesidad de introducir técnicas adicionales para fortalecer la relajación, como agregar restricciones adicionales. Esta intuición se alinea con otros trabajos teóricos y observaciones obtenidas en otros trabajos contemporáneaos para problemas relacionados como la sincronización de rotaciones [23].

Trabajo futuro

Si hay algo excitante sobre las aportaciones de esta tesis es la cantidad de oportunidades y líneas de investigación que abre.

En los últimos años se ha visto cómo problemas fundamentales en la Visión Artificial 3D, cuya resolución imponía hasta ahora una elección entre eficiencia *o* fiabilidad/garantías, pueden ser abordados a través de algoritmos que son eficientes *y* fiables. Sin embargo, creemos que nuestras contribuciones (y otras contemporáneas) tan sólo se adentran en la superficie de este campo, y queda la pregunta sobre cuántos problemas fundamentales de la Visión Artificial en 3D podemos abordar de una manera mejor, más eficiente y con garantías de optimalidad. Y también, ¿cuál puede ser el impacto de estas mejoras si se integran en las librerías estado del arte habituales que soportan el campo emergente de la percepción artificial 3D en la industria y otras aplicaciones?

A continuación, proporcionamos una lista no exhaustiva de futuras direcciones de investigación que creemos que están estrechamente conectadas con las contribuciones presentadas en esta tesis y que podrían conllevar importantes avances en la disciplina.

Certificadores rápidos y algoritmos certificables para más problemas más allá de PGO

SE-Sync [24] y Cartan-Sync [21] son dos buenos ejemplos del potencial existente en aprovechar hábilmente las relajaciones convexas ajustadas y fusionarlas con técnicas de optimización locales eficientes para proporcionar métodos óptimos, rápidos y certificables [25]. En este contexto, la capacidad de encontrar un certificado de optimalidad, de manera rápida y eficiente, es una pieza clave.

Mientras que tal certificador ha sido definido para algunos problemas (por ejemplo "Pose Graph Optimization" [18]), aún queda por encontrar certificadores rápidos similares para otros problemas para los que hemos hallado una relajación convexa ajustada, por ejemplo para el registro 3D de puntos a puntos, líneas y planos [16] o el problema de la pose relativa ("Relative Pose Problem") [17].

Relajaciones convexas ajustadas para problemas más complejos

Como se mencionó anteriormente, encontrar relajaciones ajustadas para problemas fundamentales no convexos en la Visión Artificial en 3D no es una tarea sencilla. En nuestro trabajo, abordamos problemas relevantes que, a pesar de su relativa simplicidad, aún no eran asequibles en el sentido de la optimización con garantías globales (a través de relajaciones convexas estrechas) mediante los métodos estado del arte.

Queda una miríada de problemas relevantes diferentes, para los que valdría la pena explorar la viabilidad de caracterizar relajaciones convexas ajustadas que, en caso de existir, podrían aprovecharse nuevamente para impulsar el estado del arte en métodos con garantías de optimalidad global. Estos problemas abarcan desde problemas incrementalmente más complejos, como por ejemplo el registro 3D de líneas a líneas como una alternativa geométricamente más significativa que el problema clásico de la pose relativa, basado en costes algebraicos, hasta escenarios más complejos como el registro 3D de múltiples conjuntos de primitivas (en lugar del registro por pares que tratamos aquí), y más allá con problemas que se aproximen más al problema completo de "Bundle Adjustment".

La técnica de la "escalera de Riemann" más allá de PGO

En Cartan-Sync [21] vimos cómo la aplicación del método Burer-Monteiro para producir una relajación intermedia, aunque no convexa del problema original, eventualmente condujo a la optimización en un dominio cercano al original que presentaba elevaciones ("lifts") incrementales del espacio $\mathrm{SE}(3)$, que pasaban a convertirse en una especie de pose generalizada/extendida (el grupo Cartan presentado en [21]).

A pesar de que la aparición de estas “poses generalizadas” estaba estrechamente conectada a la estructura de la formulación concreta de la sincronización de poses en Cartan-Sync [21], dado el impacto notable de esta relajación parcial en la convergencia global para el problema resultante, esto plantea la pregunta de si este “truco”, relajar parcialmente ciertos objetos no convexos como son las poses, podría aprovecharse de alguna manera fuera del contexto concreto de la formulación en Cartan-Sync e incorporarse a otros problemas en los que, incluso si no proporcionara ninguna garantía per se, podría suponer una mejora empírica en la convergencia y en la reducción de la no convexidad general del problema correspondiente.

Mejores algoritmos de inicialización

En nuestro trabajo, las relajaciones convexas propuestas se han aprovechado principalmente para la resolución global de los problemas correspondientes, dado que estas relajaciones eran ajustadas o "tight".

De manera más general, sin embargo, para otros problemas en la literatura de optimización (como por ejemplo el problema "Max-Cut" [26]) tradicionalmente las relajaciones propuestas no han sido ajustadas, pero aún así se han aprovechado para producir buenas estimaciones iniciales a través de diferentes heurísticas. Nosotros mismos propusimos un procedimiento de este tipo para los casos del problema de "Pose Graph Optimization" cuando la dualidad fuerte no se daba en [20].

De manera similar, creemos que existe gran potencial en la exploración y explotación de versiones simplificadas de las relajaciones convexas propuestas. Incluso si no son ajustadas, estas relajaciones podrían conducir a nuevos algoritmos de inicialización que tuvieran un mejor equilibrio entre la complejidad y la calidad de inicialización, lo que conduciría a métodos de optimización más efectivos (ya que la inicialización es un componente clave de cualquier sistema en la práctica, como se muestra por ejemplo en Cartan-Sync [21]).

Sobre valores atípicos ("outliers") y métodos robustos

Un caso particular pero relevante de problemas más complejos aún por tratar es el de las versiones robustas de los mismos problemas abordados aquí.

El lector familiarizado probablemente habrá notado a estas alturas la ausencia de cualquier mención a valores atípicos y métodos robustos en esta tesis, que sin duda son aspectos esenciales para cualquier método del mundo real en la práctica. Esto no es una coincidencia. Como se dijo al inicio de esta tesis, los problemas aquí abordados se caracterizaron en parte por tener como *único* principal desafío la aparición de poses en el dominio de la optimización. Esto fue intencionadamente, para que pudiéramos enfocar nuestro esfuerzo y análisis en ese único aspecto fundamental de los problemas geométricos abordados (siguiendo el principio de “divide y vencerás”).

La construcción de algoritmos certificables es mucho más complicado que simplemente encontrar soluciones empíricas. Por lo tanto, aumentar la complejidad del problema subyacente (manteniendo la certificabilidad y las garantías) generalmente requiere un buen entendimiento de cada fuente de complejidad por separado. En este sentido, se están realizando muchas investigaciones en el campo más general de la estimación robusta. Esperamos que nuestra contribución aquí llene un vacío en la comprensión y certificabilidad en el campo comparativamente más especializado de la no convexidad en torno a la estimación de poses, que es muy relevante a su vez para un área tan amplia como la Visión Artificial en 3D.

Creemos firmemente que se puede lograr un avance práctico muy significativo aunando robustez y certificabilidad en los problemas geométricos fundamentales tratados en esta tesis, y esperamos ver más investigaciones en esta dirección en el futuro de la disciplina.

Dos metodologías concretas para estimación robusta que pueden beneficiarse de ampliar nuestras contribuciones aquí son:

1. Combinar nuestra capacidad para resolver la versión no robusta de estos problemas de forma óptima, con algún método empírico para estimación robusta como la de "no-convexidad graduada" ("Graduated Non-Convexity") [27,28]. Alguno avances prometedores han ocurrido en esta dirección para algunos problemas de forma posterior a las contribuciones de esta tesis, por ejemplo en [29].

2. Caracterizar relajaciones convexas ajustadas para formulaciones robustas de los problemas que tratamos aquí, especialmente teniendo en cuenta que para muchas formulaciones típicamente robustas, por ejemplo con "estimadores-M" ("M-estimators"), el problema resultante aún se puede formular como un problema polinómico [28] (que podría presentar potencialmente relajaciones ajustadas del problema SDP correspondiente en ciertos regímenes del problema).

Part I

Introduction

Introduction

We have seen world change, both in historical records and by ourselves. World *is* changing. But world, oh my friend, world is about to change way more.

Current and future technologies

Think of all the technology involved in our everyday lives.

More than one hundred years of history of industrial mass-production endorse that automobile you drive home.

Last century also saw the rise of computers, only topped by the development of mobile computing, and concretely the smartphone. I can easily think of a few people who do not own a car. But, how easily can you think of anyone who does not have his own smartphone? What physically might appear as just a 2D surface lighting a bunch of pixels has become for the average citizen a window that lets you peak into a virtual world, the Internet, where communication and information capabilities are virtually unlimited.

If the impact of the smartphone in our daily life is apparent, we are still actively moving towards even more impactful technologies. Consider every time you drew your smartphone today, all the actions you did with it, and all the information retrieved as a result: Check the weather, use your favorite maps app to reach a destination, message your friends, your family, check the mail... Now imagine getting all that information displayed at the blink of an eye. But we can go even further: Want to see in the dark? Want to share the amazing view of the Grand Canyon with your parents, in real time, as if they were there? Need prescription glasses? Want some historic notes on that monument you are looking at? Talking to someone in a busy pub, would you like to cancel the surrounding noise? Meet Augmented Reality (see Fig. 1).

One more hit of past century, robots have proven highly valuable in the industrial environment at alleviating human labor by automating burdensome tasks, maybe even performing tasks that are out of reach for a human. We have also had the opportunity at this point to see some of these robots deployed at home. One prominent example are robot vacuum cleaners, aiming at releasing you from some household burden since 2002 (Fig. 2).

Figure 1: Some examples for Augmented Reality potential. Above: Reality as perceived today. Below: Reality as could be perceived through Augmented Reality glasses. Images taken from `https://www.oculus.com/blog/inventing-the-future/`.

Figure 2: iRobot introduced the Roomba series as far back as in 2002. Image taken from `https://www.irobot.ie/`.

Figure 3: Left: A (maybe future) humanoid robot (image taken from `https://blogs.3ds.com/northamerica/future-robots-and-ensuring-human-safety/`). Right: A drone operating in a challenging scenario (image taken from `https://newsroom.cisco.com/feature-content?type=webcontent&articleId=1940944`).

Go smarter

Now, imagine all that made *smarter*:

It has been a long day, maybe you even decided to share a well-deserved beer with colleagues after office. What about just seating in your car, relaxing, and letting it drive *you* home (rather than the other way around)!

But smart technologies have the potential to be even more impactful in our existence. If evolution has made wonders through millions of years bringing all those "magical" capabilities every single person already has to sense and communicate, Augmented Reality brings the promise to enrich our capabilities and specially our user experience interacting with the virtual world, which at this time we do mostly via our smartphones. Considering all these potential extended capabilities, want to know what is next? Imagine all these capabilities we described above being automatically served to you, based on your context, just as a natural extension of your capabilities and your will. Meet *Intelligent* Augmented Reality, meet the superhuman. Intelligent Augmented Reality is not only about being able to display virtual elements into your eyes that enhance your senses or capabilities, it is also about perceiving and understanding the world just as you do in order to take informed decisions and be proactive on how to serve you.

And what about the final bastion of a sci-fi life? If Augmented Reality has the power to enhance our cognitive capabilities, getting things *known*, so does robotics in the physical reality, getting things *done*. But we can go much beyond (Fig. 3): Robots for disaster areas, drones... They all have the potential to get things done, for us. In order for a robot to be as capable as a human, shouldn't it have, well, similar capabilities?

What does it take?

At this point, I hope the future looks exciting and promising to you, it certainly does to me. A lot of challenges exist in numerous disciplines and fields that stand in the way to reach the future we sketched above. But if there is something surprising about it all is that there are many reasons and facts to think it can be done. It is only persistence, the investment of an important amount of resources and, above all, human genius, that may bring these scenarios to become a reality.

But, in a more technical sense, what do we need to get there? As mentioned above, each and every one of these applications are multidisciplinary and require numerous tools. But here we want to bring our attention onto a fundamental need all of the technologies above share: The need for Intelligent Machine Perception.

Artificial Intelligence in on the rise. But for a machine to assist us in our daily life (in any of the fashions depicted above), it does not take only to be intelligent, but also to *perceive* our daily life just the same way we do. Only that way an automated agent may have the tools to understand the environment sufficiently well so as to know what needs to be done and when, in a proactive manner. Only that way any of the capabilities above can become part of your life and improve it in a seamless way.

If something defines the capabilities of a human being to interact and understand the environment, it is (among other things of course) its ability to perceive the 3D structure of the world we live in. This may not be the only human-like quality required by an intelligent agent to be performant in a real world, but it is for sure one fundamental pillar that enables ourselves to reason about the environment the way we do.

Luckily, the maturity reached by many of the now classical algorithms and ideas in areas such as Multiple View Geometry [1] or Optimization provide as of today a toolbox of practical 3D perception capabilities upon which many current applications are already building higher-level functionalities, and are for sure the seed for the perception capabilities needed by many of the emerging technologies referenced above.

Let us look a little bit closer into what do those fundamental 3D perception tools look like.

1 A fundamental tool of 3D Computer Vision: Bundle Adjustment

Of the many enabling techniques for 3D perception, the gold standard is probably Batch Optimization, also commonly referred to as Bundle Adjustment (BA), Structure from Motion (SfM), or Simultaneous Location and Mapping (SLAM). There are some classical distinctions between these, but in practice these boundaries have become blurrier with the advances in these fields. In very broad terms, this technique attempts to find the 3D geometry of the surrounding world, using the information provided by a sensor suite (or rig), which may range from a single monocular camera to more complex multi-camera systems, often accompanied by range sensors or Inertial Measurement Units (IMUs) (see Fig. 4). The problem intrinsically involves also finding the motion of the sensor rig, and often refining many of the internal parameters for these sensors.

This kind of techniques, with the right input data, may lead to outstanding results (see Fig. 5).

On a more technical note, state-of-the-art approaches cast this problem as an instance of Maximum-a-Posteriori (MAP) estimation, where assuming a certain probability distribution for the sensors' noise we seek for a model that maximizes the overall likelihood of all the available information. This formulation is often simplified to Maximum-Likelihood Estimation (MLE), by assuming an uniform prior on the parameters to estimate.

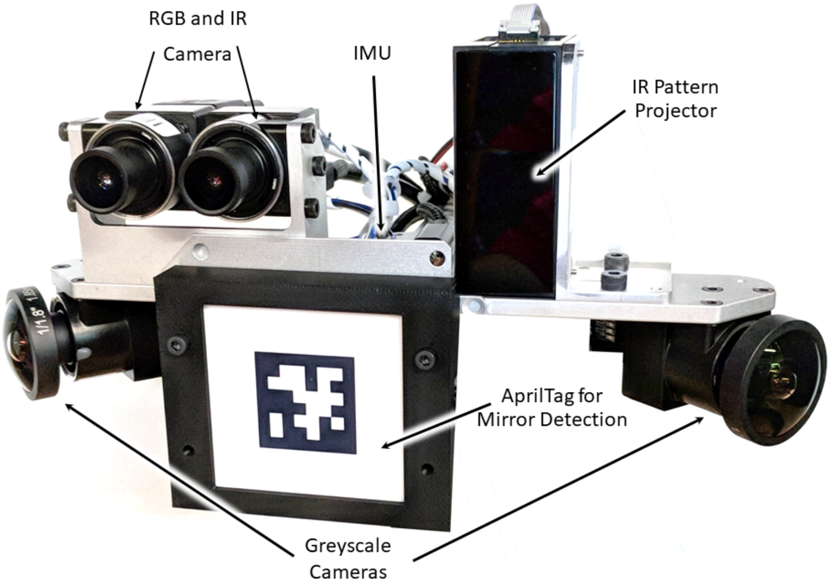

Figure 4: Professional capture rig employed in [2], consisting of greyscale, RGB and IR cameras, together with IMU and IR projector.

This is a very attractive formulation of any probabilistic problem from a theoretical point of view, given the appealing properties and guarantees that Maximum-Likelihood Estimation entails. As fancy as the above may sound, this eventually boils down to solving an optimization problem where we seek for the model parameters that maximize the consistency of all the available sensor data, or equivalently minimize the error of model predictions wrt the actual observed values in our sensors. However, this does actually entail a fundamentally hard computational task. To understand why, we should first do a quick review of optimization and its basics.

1.1 Basic notions of optimization

When solving a problem via optimization, we aim at finding the lowest possible objective or error in a certain problem region or domain.

The optimization methods typically encountered in Geometric Computer Vision and Bundle Adjustment work in an iterative manner, doing a local search from an initial point with the goal of eventually reaching a critical point (local minimum) in the objective to optimize. These methods work remarkably fast in most cases of interest, which is an essential virtue for most of the applications where 3D Computer Vision is a core component.

If the critical point found is the globally optimal one, we succeeded.

If there is *one single* critical point in the entire domain, the problem will be fundamentally easy as we just need to keep iterating in our method until eventually we reach this critical point. This is basically what defines a *convex optimization problem*: Having a single local (and thus global) minimum.

But what if multiple critical points exist? In this scenario, a typical iterative method might stagnate in the wrong local minima, preventing us to find the actually seeked solution which lies at the global minimum. This is what defines a *non-convex optimization problem*.

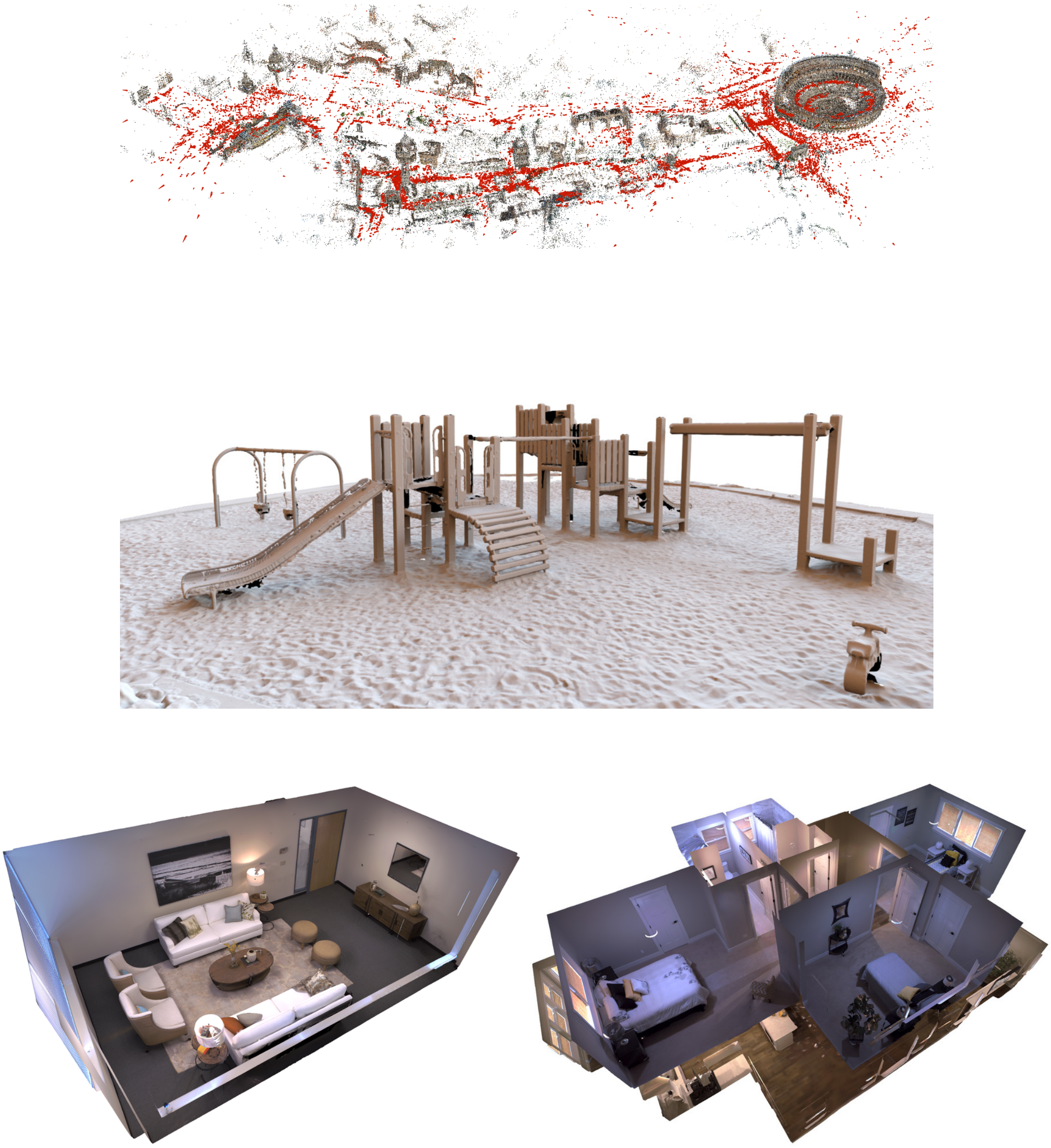

Figure 5: State-of-the-art geometric Computer Vision techniques together with the data provided by different sensors may provide outstanding 3D reconstructions. Top: SfM sparse reconstruction based on photo tourism images [3]. Middle: High-fidelity reconstruction with LIDAR technology [4]. Bottom: High-fidelity reconstructions done with professional rig in Fig. 4 [2].

As we will see later through this thesis, there may exist different sources for the non-convexity of an optimization problem: Namely either the optimization objective, the optimization domain, or both, may be non-convex, turning the overall problem non-convex as well. This is important to note, as e.g. domains involving 3D rotations and transformations are non-convex, and this is one of the critical sources of complexities this thesis revolves around.

Thus, in this scenario, the most common alternatives are either keeping the best local minimum found by a local iterative method (with no further guarantees on its global optimality whatsoever), or resorting to more involved exploratory methods (if we seek global optimality guarantees), like e.g. Branch-and-Bound, which basically does some clever tricks to recursively partition the domain and check if a partition may contain the global optimum. This latter family of methods, however, is intrinsically much more computationally expensive than the fast and scalable local iterative methods.

1.2 Back to Bundle Adjustment as an optimization problem

Now we are in a position to understand why solving Bundle Adjustment is such as hard problem. Bundle Adjustment is, by definition, a typically huge-dimensional, non-convex, non-linear optimization problem. Because of this, no existing algorithm can solve a common Bundle Adjustment instance with full optimality guarantees.

Thus, in practice, rather than finding the maximum-likelihood estimate, practical methods focus on finding a *critical point* of the underlying optimization problem, with the *hope* that this critical point will be in fact the globally optimal solution, and thus correspond to the desirable maximum-likelihood estimate. This task has been addressed by many of the currently existing solutions via applying fast first- or second-order numerical optimization methods, which is an attractive approach for the many advantages this provides in conjunction with state-of-the-art optimization techniques, such as rapid convergence speed in the proximity of a critical point or scalability in the problem size by leveraging sparsity. Overall, this has resulted in a set of tools and libraries [5, 6] that can address Bundle Adjustment in an efficient and scalable manner, and which work well under the assumption of starting *close to the optimal solution*, due to the iterative nature of these techniques.

2 Limits of usability for state-of-the-art 3D Computer Vision

Amidst the strong surge in AI-related applications, as we increase the capabilities of AI or perception-enabled automation, and thus the number of tasks we can give to them, there has been also a motivated concern as to the resilience and reliability of these systems. For many trending applications (mobile apps, etc.), failure cases may turn into a not-so-smooth user experience, which ultimately ends up with a moderately-stable-only application and a not-so-satisfied client. But for a great bunch of field-pushing applications, such as autonomous driving or robotics, a system failure (due e.g. to an algorithm's poor performance) may incur in high costs that range from damage to property to the loss of human lives [7]. Thus, in order to exploit the full potential of these techniques, we need to enable full adoption in these critical scenarios by providing an additional degree of robustness, reliability and understanding in the techniques under use.

In the context of 3D geometry (and thus in current 3D reconstruction tools such as Bundle Adjustment), a suboptimal critical point may typically be arbitrarily far from the correct solution, so finding the *actually optimal* solution to the MLE formulation turns *a necessity*. Since, as we saw, Bundle Adjustment techniques proceed mainly via iterative optimization approaches, obtaining a sufficiently good initial estimate to *bootstrap* the subsequent iterative refinement is a crucial step towards the success of the whole 3D reconstruction. For this purpose, most pipelines consist of multiple steps where the solutions to simpler subproblems are leveraged to incrementally build an initial estimate that can be fed into the final refinement scheme.

Numerous solutions have been proposed for these various classical fundamental subproblems in the context of Bundle Adjustment. Unfortunately, even these stripped down versions of Bundle Adjustment are still far from trivial. Thus, many of these core initialization algorithms that aim at securing the success of subsequent Bundle Adjustment remain still remarkably empirical in nature. This makes the whole pipeline more brittle, in the sense that even if many current solutions work perfectly in favorable circumstances, they may still fail catastrophically against a slightly not-so-favorable scenario. Ultimately, this empirical aspect of many existing solutions may render the scope of reliable operation for the algorithms more limited than we would like. Worse yet, there may be little to none certainty in which is the secure scope of operation.

Only slowly further research is being done on these fundamental building blocks of 3D Computer Vision, shedding some light onto what are the fundamental limits and what performance guarantees can we get from existing methods.

So, in conclusion, reliability is a critical feature that restricts a wider use of current 3D perception tools. Whereas reliability can (and should) be addressed in various ways (e.g. via higher sensor redundancy, performing additional checks, etc.), there is a lot to gain by improving the guarantees of current methods. Thus, in this thesis, our main contribution fall within the scope of improving the understanding and guarantees when solving several fundamental subproblems in 3D Computer Vision.

3 Some fundamental subproblems in 3D Computer Vision

Next, we will present two kinds of problems which play a fundamental role in bootstrapping Bundle Adjustment. At their core both share a common goal (and as we will see later, challenge): The computation of motion or poses. But first, what is a pose?

What is a pose?

It makes sense that in order to understand the 3D shape of the world, one essential ingredient is to describe where things are located. For a rigid object this involves two parameters: Its position and its orientation (or equivalently its rotation). The combination of position and orientation is called the *pose* of an object (less often also referred to as *attitude*). Beyond the high-level concept, in practice this description needs to be given wrt some reference coordinate system or frame (from here on the world frame).

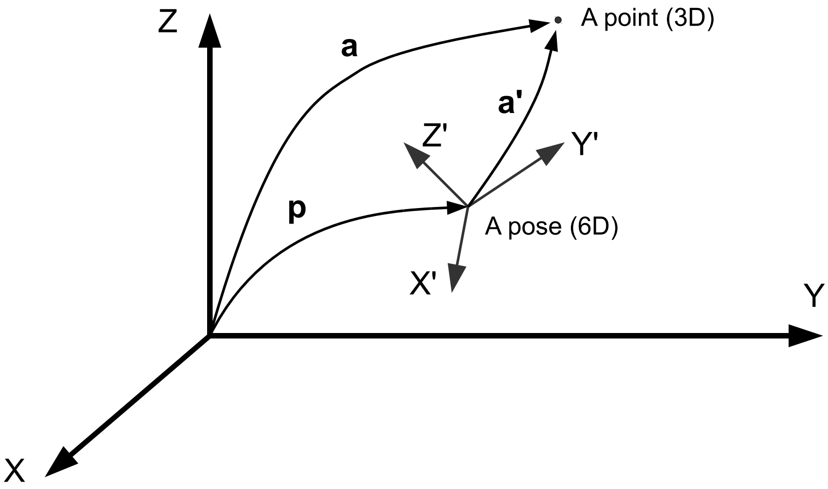

Figure 6: A pose as presented in [8]

Given a reference world frame and an object's local coordinate system (Fig. 6), the pose of the object wrt the global frame basically describes the object's frame in terms of global coordinates.

The concept of pose also underlies the idea of motion, and thus goes back to the fundamental concepts of classical mechanics, connecting with the notion of transformation of a rigid body. Thus, working with poses is also core to many other disciplines, such as robotics.

Geometric 3D Pairwise Pose-Registration problems

In registration problems, as in Bundle Adjustment, we aim at bringing together a set of observations of a common model in a consistent manner. However, in this case, it is common to introduce various simplifications that strip down part of the huge complexity in the more general Full Bundle Adjustment. Thus, whereas in Bundle Adjustment it is normal to work in complex and meaningful spaces, such as 2D points in the projected camera image, or even photometric values in the image, in registration problems it is customary to work with 3D geometric features and primitives (points, lines, planes). Also, whereas in Bundle Adjustment we often optimize over a huge bunch of parameters which include the 3D model of the environment, internal sensor parameters, etc. in registration problems we constrain ourselves to estimate the best pose to match two sets of geometric features[1], assuming the rest of parameters are good enough so that a good fit can actually happen.

Despite the simplifications above, this kind of registration problems are ubiquitous and underlie many core subproblems in classical 3D Computer Vision, playing a fundamental role to bootstrap more complex algorithms such as Bundle Adjustment. Some examples include:

- Computing the absolute pose of a visual or range sensor with respect to a given geometric model (tracking).
- Computing the relative motion of one of these sensors between two timestamps (odometry).

---

[1] Actually a registration problem might consist of more than two feature sets to fit, but for simplicity we will restrict our speech here to the base case of two sets.

- Computing the extrinsic calibration of two of these sensors in a rig (calibration).

Geometric 3D features are pervasive when working with the type of sensors above, e.g. 2D points and lines in camera images back-project into 3D lines and planes when considering projective geometry; range sensors typically detect 3D points to which we can fit 3D planes, which may intersect in turn into 3D lines. Thus, any of the above problems can be carefully cast as the registration of two sets of 3D geometric primitives (points, lines, planes).

Pose Graph Optimization

As a consequence of the ubiquity of the problems above, very often we find ourselves with redundant pairwise information regarding the relative motion or pose of various objects through time. The use of all this available redundant information to estimate a unique global model entails a mathematical problem involving only poses, which has been for a long time an object of study in various communities: Pose Graph Optimization, Motion Averaging, Pose-Synchronization... all refer to the same underlying concept and problem. Note that whereas the registration problems presented above constrained the problem complexity significantly by considering a single pose to optimize upon, in the problem of Pose Graph Optimization we may have as many as tens of thousands of poses to optimize upon (or even more). Thus, even if Pose Graph Optimization is still a strong simplification of the more general Full Bundle Adjustment or SLAM, it lies already significantly closer in terms of domain complexity and scalability.

One common underlying challenge: Optimizing over poses

Despite stripping many layers of complexity with respect to the full Bundle Adjustment problem, even in the simplest of the scenarios above, there remains a fundamental challenge: Optimizing over a pose.

From a mathematical perspective, poses (or equivalently rigid transformations) are elements of the *special Euclidean group* SE(3), one of the oldest and most studied groups in mathematics (already long before the concept of group was even invented). It might therefore come as a surprise that dealing with poses can still be a challenge in the present day, but it is!

The reason for this goes back to the basics of optimization and the difficulties implied by non-convexity. The space of 3D rotations (remember a 3D pose consists of a position and an orientation) is *non-convex*, and thus the optimization of any objective function (however simple it is) may present multiple local minima. Thus, any general optimization problem involving a pose is non-convex and thus non-trivial to solve.

As a consequence, any of the problems considered above is in general hard to solve with global optimality guarantees, although exceptions always exist. E.g. a closed-form solution based on an eigenvalue decomposition exists for the Procrustes problem between two point clouds [9, 10]. Also for some of these problems we may find all local minima using state-of-the-art approaches, but this strategy does not scale as the problem complexity increases and the number of local minima may explode.

Up until the time of the contributions in this thesis, the few existing solutions dealing with the optimization over poses in a certifiable manner were way far from the realm of real-time performance, providing optimality mainly via exploratory techniques (such as Branch-and-Bound) [11, 12].

4 Contributions and outline of this thesis

In the context of the aforementioned fundamental blocks of 3D Computer Vision, this thesis aims at characterizing different subproblems whose core difficulty stems from the estimation of poses (Sec. 3). This thesis contributes to overcoming the current reliability limitations to get guaranteed optimal solutions given the non-convex nature of those problems, by exploring practical approaches to obtain an optimal solution to relevant geometric subproblems.

The contributions of this thesis can be divided into several blocks.

4.1 Addressing different 3D Computer Vision tasks as registration problems

In the first part of this thesis, we explore the potential for geometric registration problems to model and address some typical tasks in 3D Computer Vision (Sec. 3).

One such task is the extrinsic calibration of a monocular camera and a 2D Laser Rangefinder (or LIDAR), assuming already known intrinsic calibration for both. Due to the scarce overlap on the kind of data perceived by these two sensor modalities, building the data association necessary to perform the extrinsic calibration is not straightforward. In our ICRA15 paper [13] we proposed a novel approach which overcomes this limitation by leveraging the underlying structural constraints present in any typical human-made Manhattan-like environment to establish correspondences between the 3D back-projection of extracted features in the camera image (planes and lines) and the extracted 3D features visible in the LIDAR scan (lines and points), circumventing the necessity to build a specific on-purpose calibration pattern for the task. In our IROS15 paper [14] we contributed a closed-form solution for the minimal case of the registration problem above, which basically boils down to a complete observation of an orthogonal trihedron from both the monocular camera and the 2D Laser Rangefinder.

The second tackled task is that of estimating the orientation of a monocular camera from the observation of straight lines in (again) a human-made Manhattan-like environment. In our JMIV journal [15], we provide a simple and fast solution to the problem of registering three back-projected planes in the camera frame to the corresponding 3D lines in the world when these have orthogonal directions. Then we use this solution to implement a visual gyroscope (i.e. estimate relative camera rotations from images).

Through the problems above, we confirm the main claims in the introduction of this thesis (Sec. 3): Problems involving pose estimation are very common, yet except for some specific problems where closed-form solutions can be deployed, guaranteed optimal solutions to most of the more general and interesting problems remain ellusive. That, however, we addressed in the next set of contributions.

4.2 Global optimality in 3D Geometric Registration problems

In the second block of this thesis, we revisit the fundamental topic of 3D geometric registration and show that some core classical problems of 3D Computer Vision which have been long-regarded as globally intractable (in polynomial time), are in fact solvable with optimality guarantees via an appropriate convex relaxation.

In CVPR17 [16], we show that for a quite general multi-modal 3D registration problem, involving correspondences from points to points, lines and/or planes, it is possible to get a certifiably optimal solution via an appropriate convex relaxation which turns out to be tight in virtually all tested cases.

In CVPR18 [17], we address one of the most classical and fundamental problems in Multiple View Geometry [1], the Relative Pose problem from corresponding 2D features (for calibrated cameras). Again, in this case we are able to characterize a convex relaxation of the problem that turns out to be tight in virtually all tested scenarios, providing a procedure to obtain a certifiable optimal solution in this core problem of 3D Computer Vision.

4.3 Global optimality in Pose Graph Optimization

Following with the topic of pose estimation and the inherent difficulty of related problems, in the last part of this thesis we go beyond pair-wise registration problems and focus on building a certifiably optimal solver for Pose Graph Optimization (Sec. 3). For this purpose, first, we characterize in our IROS16 paper [18] a novel convex relaxation via the Lagrangian dual of the Pose Graph Optimization problem which turns empirically tight in common circumstances. When tight, the proposed convex relaxation allows to certify the optimality of a candidate solution to the Pose Graph Optimization problem. Thanks to a more careful handling of the problem structure, the certification algorithm resulted in a much cleaner and faster procedure than for the contemporary state-of-the-art approach [19].

Building upon this same convex relaxation, in our ICRA17 paper [20] we contributed a unified approach that directly provides the certifiable globally optimal solution to a Pose Graph Optimization instance from the solution to its dual problem, if this is tight, or a competing empirical initialization otherwise in those extreme scenarios where the dual problem is not tight, e.g. when the data becomes very noisy.

Lastly, in the RAL17 journal [21] we combined the convex relaxation for Pose Graph Optimization, characterized in our previous work, with state-of-the-art advances in numerical optimization methods on manifolds to produce a fast, certifiable second-order trust-regions method that is able to converge globally to the optimal solution of a Pose Graph Optimization instance when its Lagrangian dual problem is tight.

5 Publications

The present thesis comprises the following publications and, in some cases, associated source code and supplementary videos and materials:

Journals

- J. Briales and J. Gonzalez-Jimenez, **"A Minimal Closed-form Solution for the Perspective Three Orthogonal Angles (P3oA) Problem: Application To Visual Odometry,"** in Journal of Mathematical Imaging and Vision, vol. 55, no. 3, pp. 266-283, 2015, doi: 10.1007/s10851-015-0620-x. **JCR in Computer Vision and Pattern Recognition: Q1, T1**
- J. Briales and J. Gonzalez-Jimenez, **"Cartan-Sync: Fast and Global SE(d)-Synchronization,"** in IEEE Robotics and Automation Letters (RA-L), vol. 2, no. 4, pp. 2127–2134, 2017, doi: 10.1109/LRA.2017.2718661. Not indexed yet in 2017.

International Conferences

- R. Gomez-Ojeda, J. Briales, E. Fernandez-Moral, and J. Gonzalez-Jimenez, **"Extrinsic calibration of a 2d laser-rangefinder and a camera based on scene corners,"** in IEEE International Conference on Robotics and Automation (ICRA), 2015, doi: 10.1109/ICRA.2015.7139700. **Conference Ranking: A1 (Qualis).**

  Both first and second authors had the same level of contribution.
- J. Briales and J. Gonzalez-Jimenez, **"A minimal solution for the calibration of a 2D laser-rangefinder and a camera based on scene corners,"** in IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2015, doi: 10.1109/IROS.2015.7353625. **Conference Ranking: A1 (Qualis), A (ERA).**
- J. Briales and J. González-Jiménez, **"Fast Global Optimality Verification in 3D SLAM,"** in IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2016, doi: 10.1109/IROS.2016.7759681. **Conference Ranking: A1 (Qualis), A (ERA).**
- J. Briales and J. González-Jiménez, **"Initialization of 3D Pose Graph Optimization using Lagrangian duality,"** in IEEE International Conference on Robotics and Automation (ICRA), 2017, doi: 10.1109/ICRA.2017.7989600. **Conference Ranking: A1 (Qualis).**
- J. Briales and J. González-Jiménez, **"Convex Global 3D Registration with Lagrangian Duality,"** in IEEE Conference on Computer Vision and Pattern Recognition, 2017, doi: 10.1109/CVPR.2017.595. **Conference Ranking: A1 (Qualis), A (ERA).**
- J. Briales, L. Kneip, and J. Gonzalez-Jimenez, **"A certifiably globally optimal solution to the non-minimal relative pose problem,"** in IEEE Conference on Computer Vision and Pattern Recognition, 2018, doi: 10.1109/CVPR.2018.00023. **Conference Ranking: A1 (Qualis), A (ERA).**

6 Framework and Timeline

This thesis results from five years of work by the author at the *Machine Perception and Intelligent Robotics (MAPIR)* research group[2], starting on October 2013 when I joined the MAPIR group

---

[2] http://mapir.isa.uma.es/

as a Research Assistant, already under the supervision of Prof. Javier González Jiménez. After this, I received a FPU grant (*Formación de Personal Universitario*), with this same group, supported by the Spanish Ministry of Education, which mainly funded this doctoral research. The MAPIR group is part of the Department of System Engineering and Automation, at the University of Málaga, and has extensive experience in Mobile Robotics, Computer Vision and Robotic Olfaction.

Through this period, the author completed the doctoral program in *Mechatronics Engineering* at the Department of System Engineering and Automation. In the beginning of this journey, I got to participate in the *International Computer Vision Summer School* (ICVSS), held in Sicily in 2016, which provided a stimulating venue for young researchers and Ph.D. students to meet and directly interact and discuss with world leaders in the field of Computer Vision (and I have in fact had the opportunity over the subsequent years to keep meeting many of these folks over and over, seeing the progress of the field happen with all of us and the push of time). Later on I got the opportunity to participate and present our work in multiple relevant international conferences on Robotics and Computer Vision, in Seattle (USA), Hamburg (Germany), Daejeon (South Korea), Singapore (Singapore), Honolulu (USA), Vancouver (Canada) and Salt Lake City (USA) and the International Workshop on Lines, Planes and Manhattan Models for 3-D Mapping (LPM17) celebrated jointly with the IROS conference in Vancouver, 2017.

From March 2017 to June 2017 the author completed a research stay at the *Computer Vision Group* at Chalmers University of Technology (Gothenburg, Sweeden). From January 2018 to July 2018, the author was a research intern at *Oculus (Facebook)* in Redmond (USA) under the supervision of Dr. Hauke Strasdat, and he joined again as a Full-Time Research Scientist in November 2018.

Additionally, the author has been an active reviewer of manuscripts from prestigious conferences and journals, including the *IEEE Conference on Computer Vision and Pattern Recognition* (CVPR), the *International Journal of Robotics Research* (IJRR), the *IEEE International Conference on Computer Vision* (ICCV), the *International Conference on Robotics and Automation* (ICRA), the *Pattern Recognition Letters*, the *IEEE Robotics and Automation Letters* (RA-L), the *IEEE Transactions on Robotics* (T-RO), the *International Conference on Intelligent Robots and Systems* (IROS), the *Journal of Mathematical Imaging and Vision* (JMIV) or the *Robotics: Science and Systems* (RSS).

During this thesis work, the author taught several courses as a teaching assistant at the *Escuela Técnica de Ingeniería Informática* and the *Escuela Técnica Superior de Ingeniería Industrial* at the University of Malaga.

Part II

Registration Problems in 3D Computer Vision

1

Extrinsic Calibration of a 2D Laser-Rangefinder and a Camera based on Scene Corners

Ruben Gomez-Ojeda, Jesus Briales and Javier Gonzalez-Jimenez

Published in Proc. of International Conference on Robotics and Automation (ICRA), 2015.
DOI: 10.1109/ICRA.2015.7139700

Extrinsic Calibration of a 2D Laser-Rangefinder and a Camera based on Scene Corners

Ruben Gomez-Ojeda, Jesus Briales and Javier Gonzalez-Jimenez

Abstract

Robots are often equipped with 2D laser-rangefinders (LRFs) and cameras since they complement well to each other. In order to correctly combine measurements from both sensors, it is required to know their relative pose, that is, to solve their extrinsic calibration. In this paper we present a new approach to such problem which relies on the observations of an orthogonal trihedron which is profusely found as corners in human-made scenarios. Thus, the method does not require any specific pattern, which turns the calibration process fast and simpler to perform. The estimated relative pose has proven to be also very precise since it uses two different types of constraints, *line-to-plane* and *point-to-plane*, as a result of a richer configuration than previous proposals that relies on plane or V-shaped patterns. Our approach is validated with synthetic and real experiments, showing better performance than the state-of-art methods.

2

A minimal solution for the Calibration of a 2D Laser-Rangefinder and a Camera based on Scene Corners

Jesus Briales and Javier Gonzalez-Jimenez

Published in Proc. of International Conference on Intelligent Robots and Systems (IROS), 2015.
DOI: 10.1109/IROS.2015.7353625

A minimal solution for the Calibration of a 2D Laser-Rangefinder and a Camera based on Scene Corners

Jesus Briales and Javier Gonzalez-Jimenez

Abstract

Robots are often equipped with 2D laser-rangefinders (LRFs) and cameras since they complement well to each other. In order to correctly combine the measurements from both sensors, it is required to know their relative pose, that is, to solve their extrinsic calibration. In this paper we present a simple, quick and effective minimal solution for the extrinsic calibration problem. Our approach does not require any on-purpose calibration pattern: it bases on the observation of an orthogonal trihedron, which is profusely found as corners in human-made scenarios. The proposal is validated with synthetic and real experiments, showing better performance than existing alternatives. An implementation of our approach is made available as open-source software[1].

---

[1] Code available at `http://mapir.isa.uma.es/jbriales`

3

A minimal closed-form solution for the Perspective Three orthogonal Angles (P3oA) problem. Application to visual odometry.

Jesus Briales and Javier Gonzalez-Jimenez

Published in Journal of Mathematical Imaging and Vision (JMIV), 2016.
DOI: 10.1007/s10851-015-0620-x

A minimal closed-form solution for the Perspective Three orthogonal Angles (P3oA) problem. Application to visual odometry.

Jesus Briales and Javier Gonzalez-Jimenez

Abstract

We provide a simple closed-form solution to the *Perspective Three orthogonal Angles* (P3oA) problem: given the projection of three orthogonal lines in a calibrated camera, find their 3D directions. Upon this solution, an algorithm for the estimation of the camera relative rotation between two frames is proposed. The key idea is to detect triplets of orthogonal lines in a hypothesize-and-test framework and use all of them to compute the camera rotation in a robust way. This approach is suitable for human-made environments where numerous groups of orthogonal lines exist. We evaluate the numerical stability of the P3oA solution and the performance of the odometry algorithm with synthetic and real data, comparing our results to a state-of-the-art similar approach.

Part III

Tight Convex Relaxations in 3D Registration Problems

1

Convex Global 3D Registration with Lagrangian Duality

Jesus Briales and Javier Gonzalez-Jimenez

Published in Proc. Conference on Computer Vision and Pattern Recognition (CVPR), 2017.
DOI: 10.1109/CVPR.2017.595

Convex Global 3D Registration with Lagrangian Duality

Jesus Briales and Javier Gonzalez-Jimenez

Abstract

The registration of 3D models by a Euclidean transformation is a fundamental task at the core of many application in computer vision. This problem is non-convex due to the presence of rotational constraints, making traditional local optimization methods prone to getting stuck in local minima. This paper addresses finding the globally optimal transformation in various 3D registration problems by a unified formulation that integrates common geometric registration modalities (namely *point-to-point*, *point-to-line* and *point-to-plane*). This formulation renders the optimization problem independent of both the number and nature of the correspondences.

The main novelty of our proposal is the introduction of a *strengthened* Lagrangian dual relaxation for this problem, which surpasses previous similar approaches [82] in effectiveness. In fact, even though with *no theoretical guarantees*, exhaustive empirical evaluation in both synthetic and real experiments *always* resulted on a tight relaxation that allowed to recover a guaranteed globally optimal solution by exploiting duality theory.

Thus, our approach allows for effectively solving the 3D registration with global optimality guarantees while running at a fraction of the time for the state-of-the-art alternative [83], based on a more computationally intensive Branch and Bound method.

2

A Certifiably Globally Optimal Solution to the Non-Minimal Relative Pose Problem

Jesus Briales, Laurent Kneip and Javier Gonzalez-Jimenez

Published in Proc. Conference on Computer Vision and Pattern Recognition (CVPR), 2018.
DOI: 10.1109/CVPR.2018.00023

A Certifiably Globally Optimal Solution to the Non-Minimal Relative Pose Problem

Jesus Briales, Laurent Kneip and Javier Gonzalez-Jimenez

Abstract

Finding the relative pose between two calibrated views ranks among the most fundamental geometric vision problems. It therefore appears as somewhat a surprise that a globally optimal solver that minimizes a properly defined energy over non-minimal correspondence sets and in the original space of relative transformations has yet to be discovered. This, notably, is the contribution of the present paper. We formulate the problem as a Quadratically Constrained Quadratic Program (QCQP), which can be converted into a Semidefinite Program (SDP) using Shor's convex relaxation. While a theoretical proof for the tightness of this relaxation remains open, we prove through exhaustive validation on both simulated and real experiments that our approach always finds and certifies (a-posteriori) the global optimum of the cost function.

Part IV

Tight Convex Relaxations in Pose Graph Optimization

1

Fast Global Optimality Verification in 3D SLAM

Jesus Briales and Javier Gonzalez-Jimenez

Published in Proc. International Conference on Intelligent Robots and Systems (IROS), 2016.
DOI: 10.1109/IROS.2016.7759681

Fast Global Optimality Verification in 3D SLAM

Jesus Briales and Javier Gonzalez-Jimenez

Abstract

Graph-based SLAM has proved to be one of the most effective solutions to the Simultaneous Localization and Mapping problem. This approach relies on nonlinear iterative optimization methods that in practice perform both accurately and efficiently. However, due to the non-convexity of the problem, the obtained solutions come with no guarantee of global optimality and may get stuck in local minima. The application of SLAM to many real-world applications cannot be conceived without additional control tools that detect possible suboptimalities *as soon as possible* in order to take corrective action and avoid catastrophic failure of the entire system.

This paper builds upon the state-of-the-art framework [19] in verification for this problem and introduces a novel superior formulation that leads to a much higher efficiency. While retaining the same high effectiveness, the verification times of our proposal reduce up to >50x, paving the way for faster verification in critical real applications or in embedded low-power systems. We support our claims with extensive experiments with real and simulated data.

2

Initialization of 3D Pose Graph Optimization using Lagrangian duality

Jesus Briales and Javier Gonzalez-Jimenez

Published in Proc. International Conference on Robotics and Automation (ICRA), 2017.
DOI: 10.1109/ICRA.2017.7989600

Initialization of 3D Pose Graph Optimization using Lagrangian duality

Jesus Briales and Javier Gonzalez-Jimenez

Abstract

Pose Graph Optimization (PGO) is the *de facto* choice to solve the trajectory of an agent in Simultaneous Localization and Mapping (SLAM). The Maximum Likelihood Estimation (MLE) for PGO is a non-convex problem for which no known technique is able to guarantee a globally optimal solution under general conditions. In recent years, Lagrangian duality has proved suitable to provide good, frequently *tight* relaxations of the hard PGO problem through *convex* Semidefinite Programming (SDP). In this work, we build from the state-of-the-art Lagrangian relaxation [111] and contribute a complete *recovery procedure* that, given the (tractable) optimal solution of the relaxation, provides either the *optimal* MLE solution if the relaxation is tight, or a remarkably good feasible guess if the relaxation is non-tight, which occurs in specially challenging PGO problems (very noisy observations, low graph connectivity, etc.). In the latter case, when used for initialization of local iterative methods, our approach outperforms other state-of-the-art approaches converging to better solutions. We support our claims with extensive experiments.

3

Cartan-Sync: Fast and Global SE(d)-Synchronization

Jesus Briales and Javier Gonzalez-Jimenez

Published in Robotics and Automation Letters (RAL), 2017.
DOI: 10.1109/LRA.2017.2718661

Cartan-Sync: Fast and Global SE(d)-Synchronization

Jesus Briales and Javier Gonzalez-Jimenez

Abstract

This work addresses the fundamental problem of Pose Graph Optimization (PGO), which is pervasive in the context of SLAM, and widely known as SE(d)-Synchronization in the mathematical community. Our contribution is twofold. First, we provide a novel, elegant and compact matrix formulation of the Maximum Likelihood Estimation (MLE) for this problem, drawing interesting connections with the Connection Laplacian of a graph object. Secondly, even though the MLE problem is non-convex and computationally intractable in general, we exploit recent advances in convex relaxations of PGO and Riemannian techniques for low-rank optimization to yield an *a-posteriori certifiably globally optimal* algorithm [144] that is also *fast* and *scalable*.

This work builds upon a fairly demanding mathematical machinery, but beyond the theoretical basis presented, we demonstrate its performance through extensive experimentation in common large-scale SLAM datasets. The proposed framework, `Cartan-Sync`, is up to one order of magnitude faster that the state-of-the-art `SE-Sync` [24] in some important scenarios (e.g. the KITTI dataset).

We make the code for `Cartan-Sync` available at bitbucket.org/jesusbriales/cartan-sync, along with some examples and guides for a friendly use by researchers in the field, hoping to promote further adoption and exploitation of these techniques in the robotics community.

Part V

Conclusions

Conclusions

In this thesis we have addressed multiple fundamental (optimization) problems in 3D Computer Vision which, despite the relative maturity of the field, remained inherently hard to solve due to the non-convex nature of optimization over 3D poses.

First, we explored and showcased the ubiquity of 3D registration problems in multiple forms (involving the registration of different geometric primitives), and how these serve as the core estimation problem for solving different tasks in 3D Computer Vision, such as calibration or odometry (Sec. II).

Then, we turned onto the task of solving with global optimality guarantees *and* efficiently some of these generic or ubiquitous 3D registration problems (Sec. III), as well as the related or complementary problem of synchronizing/averaging multiple global poses based on relative observations between them (Sec. IV), which naturally arises in many applications when combining the output of multiple 3D registration problems. Whereas different approaches or techniques exist to solve optimization problems with global optimality guarantees, e.g. exploratory techniques like Branch-and-Bound (BnB) or algebraic techniques based on the computation of all critical points for the cost function, characterizing a tight convex relaxation stands out as one of the most effective and scalable approaches. In particular, the contributions in this thesis showed that solving the Semidefinite Programs (SDP) arising from the exercised relaxations for the different problems beat alternatives based on BnB for generic global optimization over poses, even if using just off-the-shelf solvers.

Furthermore, we got to show how by cleverly leveraging the existence of these tight relaxations we may even be able to deploy tailored certifiable algorithms [21] that compete with state-of-the-art local iterative solvers in terms of performance while still additionally providing the global optimality guarantees only possible through the corresponding tight convex relaxation.

On the art of characterizing tight relaxations

Finding a tight convex relaxation for a given problem (the key piece for most of our contributions) is, of course, not a trivial job. Any type or family of non-convex problems, and in particular any Quadratically Constrained Quadratic Program (QCQP), remains fundamentally hard (in Computer Science slang, NP-hard) to solve with global optimality guarantees, until we find a specifically crafted approach that proves otherwise (either with theoretical or empirical guarantees).

In the approach followed in this thesis for finding tight convex relaxations to some of these problems, we consistently relied on the fact that different equivalent formulations for a given optimization problem can lead to significantly different relaxations, ranging from useless to even *tight* [22]. While this aspect of optimization problems turns the crafting of relaxed alternatives into some kind of mathematical *art*, it is also true (and showcased in this thesis) that certain tricks and methods work well across different problems, enabling us to explore potential valuable relaxations in more systematic ways. If we had to highlight one such trick from between those leveraged in this thesis, that would be without any doubt the high potential of introducing additional equivalent but independent constraints into a problem constraint set. Such trick has been instrumental in attaining the successful, tight relaxations encountered for the addressed 3D registration problems (Sec. II).

On the power of data

Yet another relevant observation from our work through this thesis is how more data can make optimization problems easier (yes, data is not only powerful for Machine Learning). More data, meaning a larger number of correspondences in registration problems, or higher graph connectivity in synchronization problems, can make these optimization problems inherently simpler. This is reflected by the fact that simpler relaxations of those problem instances become tight and enable to still solve the problem optimally. This is our intuition after the fact that, even if our proposed relaxations for 3D registration problems stand out for its ability to remain (empirically) tight in all tested problem instances, simpler relaxations can still often deal with those problem instances with a large number of observations. It is also the intuition behind the fact that for Pose Synchronization, despite the large dimensionality and overall complexity of the problem domain, a relatively straightforward relaxation of the problem turns out to be tight in practice, without the need to involve any relaxation strengthening tricks like adding additional constraints. This intuition aligns with other theoretical work and insights done in contemporary research for close problems like Rotation Synchronization [23].

Future work

If there is something we are most excited about the contributions in this thesis, it is the number of opportunities and lines of research it opens up.

The last few years have seen how many fundamental problems in 3D Computer Vision, whose resolution imposed so far a choice between efficiency *or* reliability/guarantees, might be actually amenable via algorithms that are both efficient *and* reliable. Yet, we feel our (and other contemporary) contributions have just scratched the surface of it, and there remains the research question as to how many fundamental problems of 3D Computer Vision we can tackle in better, more efficient and guaranteed ways. And also, what can be the impact of such improvements if these make their way to the core of the state-of-the-art toolboxes that support the emerging field of 3D Machine Perception in industry and applications?

Next we provide a non-comprehensive list of future directions of research that we feel are tightly connected with the contributions presented in this thesis, and that might hold potential breakthroughs for the field.

Fast certifiers and certifiable algorithms for more problems beyond PGO

SE-Sync [24] and Cartan-Sync [21] are two good examples of the potential for cleverly leveraging tight convex relaxations and fusing these with efficient, local optimization frameworks to provide fast and certifiably optimal solvers [25]. In this context, the ability to compute an optimality certificate, fast and efficiently, is a key piece for the whole fast, certifiable paradigm.

Whereas such an efficient optimality certifier has been found for some problems (e.g. for Pose Graph Optimization [18]), it still remains to find similar fast optimality certifiers for other problems for which we characterized a tight convex relaxation, e.g. 3D registration of points to points, lines and planes [16] or the Relative Pose problem [17].

Tight convex relaxations for more complex problems

As mentioned earlier, finding tight relaxations for fundamental non-convex problems in 3D Computer Vision is not a straighforward task. In our research, we took upon relevant problems which despite their relative simplicity were still non-manageable in the globally optimal sense (via tight convex relaxations) by state-of-the-art methods.

There remains a myriad of different relevant problems, for which it would be worth to explore the viability of characterizing tight convex relaxations that, in case of existing, might again be leveraged to push the state-of-the-art in solvers with guarantees. These problems span from incrementally more complex problems, like e.g. 3D lines-to-lines registration as a more geometrically meaningful alternative to the algebraic cost of the calibrated Relative Pose problem, to more complex scenarios like 3D registration of multiple sets of primitives (rather than pair-wise registration we dealt with here), and all the way up to closer approximations of the full, gold Bundle Adjustment problem.

The Riemannian staircase trick beyond PGO

In Cartan-Sync [21] we saw how the application of the Burer-Monteiro approach to produce an intermediate, yet non-convex relaxation of the original problem eventually led to optimizing on a close sibling of the original domain which featured incremental lifts of the $\mathrm{SE}(3)$ space into some kind of generalized/extended pose (the Cartan group featured in [21]).

Even though the appearance of these "lifted poses" was tightly connected to the concrete structure of the concrete formulation of Pose Synchronization in Cartan-Sync [21], given the remarkable impact of this partial relaxation on the global convergence capabilities for the resulting problem, we cannot but wonder if this "trick", partially lifting certain non-convex objects like poses, could be leveraged somehow outside the concrete structure of the Cartan-Sync formulation and baked in into other problems where, even if it would not provide any guarantees per se, it could suppose an empirical improvement in practice on convergence capabilities and in reducing the overall non-convexity of the corresponding problem.

Better initialization algorithms

In our work, the proposed convex relaxations have been exploited mainly towards the global resolution of the corresponding problems, given the tightness of these relaxations.

More generally, though, for other problems in the optimization literature (a gold standard example being the Max-Cut problem [26]) proposed relaxations have traditionally not been tight but still leveraged for producing good initial estimates via different heuristics. We proposed such a procedure ourselves for Pose Graph Optimization cases where strong duality would not hold in [20].

In a similar fashion, we think there is potential in exploring and exploiting simplified versions of the proposed convex relaxations. Even if not tight, these might still lead to novel initialization algorithms which had a better trade-off between complexity and initialization quality, thus leading to more effective optimization pipelines (as initialization is a key component of any practical solver, as motivated e.g. in Cartan-Sync [21]).

On outliers and robust methods

A particular yet relevant case of more complex problems yet-to-be-dealt-with is that of robust versions of the same problems tackled here.

The acquainted reader might have probably noticed at this point the absence of any mention to outliers and robust methods, which are undoubtedly essential aspects to any real-world method in practice, when exploring convex relaxations in Parts III and IV. This is not coincidental. As stated at the beginning of this thesis, the problems addressed here were characterized in part for having as its main challenge the appearance of poses in the optimization domain, only. This was intentionally so, so that we could focus our effort and analysis on that single fundamental aspect ("divide and rule") of the addressed geometric problems.

Building certifiable algorithms is significantly more involved than just finding empirical solutions. Thus, increasing in the complexity of the underlying problem (while keeping certifiability and guarantees) typically requires a good understanding of each source of complexity on its own. In this sense, much research is being done in the more general field of robust estimation. It is our hope that our contribution here fills a gap of understanding and certifiability in the comparatively more specialized field of non-convexity around pose estimation, which is yet vastly important for a general enough topic such as 3D Computer Vision.

It is our firm belief that very valuable progress can be done in joining robustness and certifiability in the fundamental geometric problems treated in this thesis, and we expect to see further research in this direction by follow-up works.

Two concrete robust methodologies which may benefit from or extend our contributions here are:

1. Combine our ability to solve the non-robust version of these problems optimally, with some empirical framework for robust estimation like Graduated Non-Convexity [27, 28]. Some promising work has been done in this direction for some problems after the contributions of this thesis, e.g. in [29].

2. Characterize tight convex relaxations for robust formulations of the problems we dealt with here, specially since for many typically robust formulations, e.g. with M-estimators, the resulting problem can be still formulated as a polynomial problem [28] (which might again potentially feature tight SDP relaxations in certain problem regimes).

Bibliography

[1] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. New York, NY, USA: Cambridge University Press, 2003.

[2] J. Straub, T. Whelan, L. Ma, Y. Chen, E. Wijmans, S. Green, J. J. Engel, R. Mur-Artal, C. Ren, S. Verma, A. Clarkson, M. Yan, B. Budge, Y. Yan, X. Pan, J. Yon, Y. Zou, K. Leon, N. Carter, J. Briales, T. Gillingham, E. Mueggler, L. Pesqueira, M. Savva, D. Batra, H. M. Strasdat, R. De Nardi, M. Goesele, S. Lovegrove, and R. Newcombe, "The Replica Dataset: A Digital Replica of Indoor Spaces," *arXiv:1906.05797 [cs, eess]*, Jun. 2019.

[3] J. L. Schonberger and J.-M. Frahm, "Structure-from-motion revisited," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 4104–4113.

[4] A. Knapitsch, J. Park, Q.-Y. Zhou, and V. Koltun, "Tanks and temples: Benchmarking large-scale scene reconstruction," *ACM Transactions on Graphics (ToG)*, vol. 36, no. 4, pp. 1–13, 2017.

[5] R. Kummerle, G. Grisetti, H. Strasdat, K. Konolige, and W. Burgard, "G2o: A general framework for graph optimization," in *Proc. Int. Conf. Robot. and Autom. (ICRA)*. Shanghai: IEEE, May 2011, pp. 3607–3613.

[6] S. Agarwal and K. Mierle, "Ceres solver: Tutorial & reference," *Google Inc*, vol. 2, p. 72, 2012.

[7] B. Vlasic and N. E. Boudette, "As U.S. investigates fatal Tesla crash, company defends Autopilot system," *The New York Times*, 2016.

[8] J. Blanco, "A tutorial on SE(3) transformation parameterizations and on-manifold optimization," University of Malaga, Tech. Rep., 2010.

[9] K. S. Arun, T. S. Huang, and S. D. Blostein, "Least-squares fitting of two 3-D point sets," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 9, no. 5, pp. 698–700, 1987.

[10] B. K. P. Horn, "Closed-form solution of absolute orientation using unit quaternions," *Journal of the optical society of America*, vol. 4, no. 4, pp. 629–642, 1987.

[11] R. Hartley and F. Kahl, "Optimal algorithms in multiview geometry," in *Asian Conference on Computer Vision*. Springer, 2007, pp. 13–34.

[12] F. Kahl, S. Agarwal, M. K. Chandraker, D. Kriegman, and S. Belongie, "Practical Global Optimization for Multiview Geometry," *International Journal of Computer Vision*, vol. 79, no. 3, pp. 271–284, Sep. 2008.

[13] R. Gomez-Ojeda, J. Briales, E. Fernandez-Moral, and J. Gonzalez-Jimenez, "Extrinsic calibration of a 2d laser-rangefinder and a camera based on scene corners," in *Proceedings - IEEE International Conference on Robotics and Automation*, 2015, pp. 3611–3616.

[14] J. Briales and J. Gonzalez-Jimenez, "A minimal solution for the calibration of a 2D laser-rangefinder and a camera based on scene corners," in *Proc. Int. Conf. Intell. Robots Syst. (IROS)*, Sep. 2015, pp. 1891–1896.

[15] ——, "A Minimal Closed-form Solution for the Perspective Three Orthogonal Angles (P3oA) Problem: Application To Visual Odometry," *Journal of Mathematical Imaging and Vision*, pp. 1–18, 2015.

[16] J. Briales and J. González-Jiménez, "Convex Global 3D Registration with Lagrangian Duality," in *International Conference on Computer Vision and Pattern Recognition*, Jul. 2017.

[17] J. Briales, L. Kneip, and J. Gonzalez-Jimenez, "A certifiably globally optimal solution to the non-minimal relative pose problem," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 145–154.

[18] J. Briales and J. González-Jiménez, "Fast Global Optimality Verification in 3D SLAM," in *International Conference on Intelligent Robots and Systems (IROS)*. IEEE/RSJ, 2016.

[19] L. Carlone, D. M. Rosen, G. Calafiore, J. J. Leonard, and F. Dellaert, "Lagrangian duality in 3D SLAM: Verification techniques and optimal solutions," in *Proc. Int. Conf. Intell. Robots Syst. (IROS)*, Hamburg, Sep. 2015, pp. 125–132.

[20] J. Briales and J. González-Jiménez, "Initialization of 3D Pose Graph Optimization using Lagrangian duality," in *Int. Conf. on Robotics and Automation (ICRA)*. IEEE, 2017.

[21] J. Briales and J. Gonzalez-Jimenez, "Cartan-Sync: Fast and Global SE(d)-Synchronization," *IEEE Robotics and Automation Letters*, vol. 2, no. 4, pp. 2127–2134, 2017.

[22] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[23] A. Eriksson, C. Olsson, F. Kahl, and T.-J. Chin, "Rotation averaging and strong duality," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 127–135.

[24] D. M. Rosen, L. Carlone, A. S. Bandeira, and J. J. Leonard, “SE-Sync: A Certifiably Correct Algorithm for Synchronization over the Special Euclidean Group,” *arXiv preprint arXiv:1612.07386*, pp. 1–49, 2016.

[25] A. S. Bandeira, “A note on Probably Certifiably Correct algorithms,” *arXiv preprint arXiv:1509.00824*, 2015.

[26] M. F. Anjos and H. Wolkowicz, “Strengthened semidefinite relaxations via a second lifting for the max-cut problem,” *Discrete Applied Mathematics*, vol. 119, no. 1, pp. 79–106, 2002.

[27] Q.-Y. Zhou, J. Park, and V. Koltun, “Fast Global Registration,” in *European Conference on Computer Vision*. Springer, 2016, pp. 766–782.

[28] M. J. Black and A. Rangarajan, “On the Unification of Line Processes, Outlier Rejection, and Robust Statistics with Applications in Early Vision,” *International journal of computer vision*, vol. 19, no. 1, pp. 57–91, 1996.

[29] H. Yang, P. Antonante, V. Tzoumas, and L. Carlone, “Graduated Non-Convexity for Robust Spatial Perception: From Non-Minimal Solvers to Global Outlier Rejection,” *arXiv:1909.08605 [cs, math]*, Sep. 2019.

[30] C. Früh and A. Zakhor, “An automated method for large-scale, ground-based city model acquisition,” *International Journal of Computer Vision*, vol. 60, no. 1, pp. 5–24, 2004.

[31] J.-L. Blanco, F.-A. Moreno, and J. Gonzalez, “A collection of outdoor robotic datasets with centimeter-accuracy ground truth,” *Autonomous Robots*, vol. 27, no. 4, pp. 327–351, 2009.

[32] C. Premebida and U. J. C. Nunes, “Fusing LIDAR, camera and semantic information: A context-based approach for pedestrian detection,” *The International Journal of Robotics Research*, p. 0278364912470012, 2013.

[33] B. Douillard, D. Fox, F. Ramos, and H. Durrant-Whyte, “Classification and semantic mapping of urban environments,” *The international journal of robotics research*, vol. 30, no. 1, pp. 5–32, 2011.

[34] E. Fernández-Moral, J. G. Jiménez, P. Rives, and V. Arévalo, “Extrinsic calibration of a set of range cameras in 5 seconds without pattern,” in *2014 {IEEE/RSJ} International Conference on Intelligent Robots and Systems, Chicago, IL, USA, September 14-18, 2014*, 2014, pp. 429–435.

[35] S. Wasielewski and O. Strauss, “Calibration of a multi-sensor system laser rangefinder/camera,” in *Proceedings of the Intelligent Vehicles' 95. Symposium*. IEEE, 1995, pp. 472–477.

[36] K. Kwak, D. F. Huber, H. Badino, and T. Kanade, “Extrinsic calibration of a single line scanning lidar and a camera,” in *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*. Ieee, Sep. 2011, pp. 3283–3289.

[37] Q. Zhang and R. Pless, “Extrinsic calibration of a camera and laser range finder (improves camera calibration),” in *Proc. Int. Conf. Intell. Robots Syst. (IROS)*. IEEE, 2004, pp. 2301–2306.

[38] Z. Zhang, “Flexible camera calibration by viewing a plane from unknown orientations,” in *Proceedings of the Seventh Ieee International Conference on Computer Vision*, vol. 1. Ieee, 1999, pp. 666–673.

[39] F. Vasconcelos, J. P. Barreto, and U. Nunes, “A minimal solution for the extrinsic calibration of a camera and a laser-rangefinder.” *IEEE transactions on pattern analysis and machine intelligence*, vol. 34, no. 11, pp. 2097–107, Nov. 2012.

[40] L. Zhou, “A New Minimal Solution for the Extrinsic Calibration of a 2D LIDAR and a Camera Using Three Plane-Line Correspondences,” *IEEE Sensors Journal*, vol. 14, no. 2, pp. 442–454, Feb. 2014.

[41] R. G. von Gioi, J. Jakubowicz, J.-M. Morel, and G. Randall, “{LSD:} {A} Fast Line Segment Detector with a False Detection Control,” *{IEEE} Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 4, pp. 722–732, 2010.

[42] V. Nguyen, A. Martinelli, N. Tomatis, and R. Siegwart, “A comparison of line extraction algorithms using 2D laser rangefinder for indoor mobile robotics,” in *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2005, pp. 1929–1934.

[43] J. Clarke, “Modelling uncertainty: A primer,” Department of Engineering Science, Oxford University, Tech. Rep., 1998.

[44] O. Naroditsky, A. Patterson, and K. Daniilidis, “Automatic alignment of a camera with a line scan LIDAR system,” in *2011 IEEE International Conference on Robotics and Automation*. Ieee, May 2011, pp. 3429–3434.

[45] H. Yang, X. Liu, and I. Patras, “A simple and effective extrinsic calibration method of a camera and a single line scanning lidar,” in *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*. IEEE, 2012, pp. 1439–1442.

[46] J. Zhang, M. Kaess, and S. Singh, “Real-time depth enhanced monocular odometry,” in *Proc. Int. Conf. Intell. Robots Syst. (IROS)*, Sep. 2014, pp. 4973–4980.

[47] C. Bodensteiner, W. Hubner, K. Jungling, P. Solbrig, and M. Arens, “Monocular Camera Trajectory Optimization using LiDAR data,” *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, pp. 2018–2025, Nov. 2011.

[48] P. Tulsuk, P. Srestasathiern, M. Ruchanurucks, T. Phatrapornnant, and H. Nagahashi, “A novel method for extrinsic parameters estimation between a single-line scan LiDAR and a camera,” in *Intelligent Vehicles Symposium Proceedings, 2014 IEEE*, 2014, pp. 781–786.

[49] M. Haralick, Robert M. and Lee, Chung-Nan and Ottenberg, Karsten and Nölle, "Review and Analysis of Solutions of the Three Point Perspective Pose Estimation Problem," *Int. J. Comput. Vision*, vol. 356, no. 1994, pp. 331–356, 1994.

[50] Y. Bok, D.-g. Choi, P. Vasseur, and I. S. Kweon, "Extrinsic Calibration of Non-overlapping Camera-Laser System using Structured Environment," in *Proc. Int. Conf. Intell. Robots Syst. (IROS)*, 2014.

[51] E. Fernández-Moral, V. Arévalo, and J. González-Jiménez, "Extrinsic calibration of a set of 2D laser rangefinders," in *International Conference on Robotics and Automation (ICRA 2015)*. IEEE, 2015.

[52] K. Kanatani, "Constraints on length and angle," *Computer Vision, Graphics, and Image Processing*, vol. 41, pp. 28–42, 1988.

[53] B. M. Haralick, C.-N. Lee, K. Ottenberg, and M. Nölle, "Review and analysis of solutions of the three point perspective pose estimation problem," *International journal of computer vision*, vol. 13, no. 3, pp. 331–356, 1994.

[54] V. Lepetit, F. Moreno-Noguer, and P. Fua, "EPnP: An Accurate O(n) Solution to the PnP Problem," *International Journal of Computer Vision*, vol. 81, no. 2, pp. 155–166, Jul. 2008.

[55] J. A. Hesch and S. I. Roumeliotis, "A direct least-squares (DLS) method for PnP," in *IEEE International Conference on Computer Vision (ICCV), 2011*. IEEE, 2011, pp. 383–390.

[56] S. Li, C. Xu, and M. Xie, "A robust O (n) solution to the perspective-n-point problem," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 34, no. 7, pp. 1444–1450, 2012.

[57] F. M. Mirzaei and S. I. Roumeliotis, "Globally optimal pose estimation from line correspondences," in *2011 IEEE International Conference on Robotics and Automation*. Ieee, May 2011, pp. 5581–5588.

[58] L. Zhang, C. Xu, K. Lee, and R. Koch, "Robust and efficient pose estimation from line correspondences," in *Computer Vision–ACCV 2012*, 2013, pp. 1–14.

[59] R. Hartley, "Lines and points in three views and the trifocal tensor," *International Journal of Computer Vision*, 1997.

[60] A. Bartoli and P. Sturm, "Structure-from-motion using lines: Representation, triangulation, and bundle adjustment," *Computer Vision and Image Understanding*, 2005.

[61] J. Košecká and W. Zhang, "Extraction, matching, and pose recovery based on dominant rectangular structures," *Computer Vision and Image Understanding*, vol. 100, no. 3, pp. 274–293, Dec. 2005.

[62] G. Schindler, "Line-based structure from motion for urban environments," in *3D Data Processing, Visualization, and Transmission, Third International Symposium On*, 2006, pp. 846–853.

[63] W. Förstner, "Optimal vanishing point detection and rotation estimation of single images from a legoland scene," in *Proc. ISPRS Commission III Symp. Photogramm. Comput. Vis. Image Anal*, vol. XXXVIII, 2010, pp. 157–162.

[64] F. M. Mirzaei and S. I. Roumeliotis, "Optimal estimation of vanishing points in a Manhattan world," in *Proceedings of the IEEE International Conference on Computer Vision*, 2011, pp. 2454–2461.

[65] J. C. Bazin, Y. Seo, C. Demonceaux, P. Vasseur, K. Ikeuchi, I. Kweon, and M. Pollefeys, "Globally optimal line clustering and vanishing point estimation in Manhattan world," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference On*. IEEE, 2012, pp. 638–645.

[66] A. Elqursh and A. Elgammal, "Line-based relative pose estimation," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference On*. Ieee, Jun. 2011, pp. 3049–3056.

[67] M. Finotto and E. Menegatti, "Humanoid gait stabilization based on omnidirectional visual gyroscope," in *Workshop on Humanoid Soccer Robots (Humanoids' 09)*. Citeseer, 2009.

[68] E. Rondon, L. Carrillo, and I. Fantoni, "Vision-based altitude, position and speed regulation of a quadrotor rotorcraft." in *Proc. Int. Conf. Intell. Robots Syst. (IROS)*. Ieee, 2010, pp. 628–633.

[69] S. Barnard, "Choosing a basis for perceptual space," *Computer vision, graphics, and image processing*, 1985.

[70] Y. Wu, S. S. Iyengar, R. Jain, and S. Bose, "A new generalized computational framework for finding object orientation using perspective trihedral angle constraint," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 10, pp. 961–975, 1994.

[71] M. Dhome, M. Richetin, J.-T. Lapreste, and G. Rives, "Determination of the attitude of 3D objects from a single perspective view," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 12, pp. 1265–1278, 1989.

[72] A. Criminisi, I. Reid, and A. Zisserman, "Single view metrology," *International Journal of Computer Vision*, 2000.

[73] M. E. Antone and S. Teller, "Automatic recovery of relative camera rotations for urban scenes," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2000.*, vol. 2. IEEE, 2000, pp. 282–289.

[74] J. Košecká and W. Zhang, "Video compass," in *Computer Vision—ECCV 2002*, 2002, pp. 476–490.

[75] P. Denis, J. H. Elder, and F. J. Estrada, "Efficient edge-based methods for estimating manhattan frames in urban imagery," in *European Conference on Computer Vision*. Springer, 2008, pp. 197–210.

[76] J. C. Bazin and M. Pollefeys, "3-line RANSAC for orthogonal vanishing point detection," in *IEEE International Conference on Intelligent Robots and Systems*, 2012, pp. 4282–4287.

[77] D. Nistér, O. Naroditsky, and J. Bergen, "Visual odometry for ground vehicle applications," *Journal of Field Robotics*, vol. 23, no. 1, pp. 3–20, 2006.

[78] R. Hartley, J. Trumpf, Y. Dai, and H. Li, "Rotation Averaging," *International Journal of Computer Vision*, vol. 103, no. 3, pp. 267–305, Jan. 2013.

[79] J. C. Gower and G. B. Dijksterhuis, *Procrustes Problems*. Oxford University Press Oxford, 2004, vol. 3.

[80] G. H. Golub and C. F. Van Loan, *Matrix Computations (3rd Ed.)*. Baltimore, MD, USA: Johns Hopkins University Press, 1996.

[81] J. Sturm, N. Engelhard, F. Endres, W. Burgard, and D. Cremers, "A benchmark for the evaluation of RGB-D SLAM systems," in *Proc. Int. Conf. Intell. Robots Syst. (IROS)*. IEEE, 2012, pp. 573–580.

[82] C. Olsson and A. Eriksson, "Solving quadratically constrained geometrical problems using lagrangian duality," in *Pattern Recognition, 2008. ICPR 2008. 19th International Conference On*. IEEE, 2008, pp. 1–5.

[83] C. Olsson, F. Kahl, and M. Oskarsson, "Branch-and-Bound Methods for Euclidean Registration Problems," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 5, pp. 783–794, 2009.

[84] U. Castellani and A. Bartoli, "3d shape registration," in *3D Imaging, Analysis and Applications*. Springer, 2012, pp. 221–264.

[85] F. Pomerleau, F. Colas, and R. Siegwart, "A review of point cloud registration algorithms for mobile robotics," *Foundations and Trends in Robotics (FnTROB)*, vol. 4, no. 1, pp. 1–104, 2015.

[86] P. J. Besl and N. D. McKay, "Method for registration of 3-D shapes," in *Robotics-DL Tentative*. International Society for Optics and Photonics, 1992, pp. 586–606.

[87] Y. Chen and G. Medioni, "Object modelling by registration of multiple range images," *Image and vision computing*, vol. 10, no. 3, pp. 145–155, 1992.

[88] Y. Taguchi, Y.-D. Jian, S. Ramalingam, and C. Feng, "Point-plane SLAM for hand-held 3D sensors," in *Robotics and Automation (ICRA), 2013 IEEE International Conference On*. IEEE, 2013, pp. 5182–5189.

[89] L. Ma, C. Kerl, J. Stückler, and D. Cremers, “CPA-SLAM: Consistent plane-model alignment for direct RGB-D SLAM,” in *2016 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2016, pp. 1285–1291.

[90] B. K. P. Horn, “Closed-form solution of absolute orientation using unit quaternions,” *Journal of the optical society of America*, vol. 4, no. 4, pp. 629–642, 1987.

[91] C. Olsson, F. Kahl, and M. Oskarsson, “The registration problem revisited: Optimal solutions from points, lines and planes,” in *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference On*, vol. 1. IEEE, 2006, pp. 1206–1213.

[92] H. Li and R. Hartley, “The 3D-3D registration problem revisited,” in *2007 IEEE 11th International Conference on Computer Vision*. IEEE, 2007, pp. 1–8.

[93] C. Papazov and D. Burschka, “Stochastic global optimization for robust point set registration,” *Computer Vision and Image Understanding*, vol. 115, no. 12, pp. 1598–1609, 2011.

[94] J. Yang, H. Li, and Y. Jia, “Go-icp: Solving 3d registration efficiently and globally optimally,” in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 1457–1464.

[95] A. Segal, D. Haehnel, and S. Thrun, “Generalized-ICP.” in *Robotics: Science and Systems*, vol. 2, 2009.

[96] J. Serafin and G. Grisetti, “NICP : Dense Normal Based Point Cloud Registration,” *Preprint submitted to 2015 IEEE International Conference on Robotics and Automation*, p. 8, 2015.

[97] E. Fernández-Moral, J. González-Jiménez, and V. Arévalo, “Extrinsic calibration of 2D laser rangefinders from perpendicular plane observations,” *The International Journal of Robotics Research*, vol. 34, no. 11, pp. 1401–1417, 2015.

[98] E. Fernández-Moral, W. Mayol-Cuevas, V. Arévalo, and J. González-Jiménez, “Fast place recognition with plane-based maps,” in *Robotics and Automation (ICRA), 2013 IEEE International Conference On*. IEEE, 2013, pp. 2719–2724.

[99] R. F. Salas-Moreno, B. Glocken, P. H. J. Kelly, and A. J. Davison, “Dense planar slam,” in *Mixed and Augmented Reality (ISMAR), 2014 IEEE International Symposium On*. IEEE, 2014, pp. 157–164.

[100] R. Gomez-ojeda, J. Briales, and J. Gonzalez-jimenez, “PL-SVO: Semi-Direct Monocular Visual Odometry by Combining Points and Line Segments,” in *IEEE International Conference on Intelligent Robots and Systems*, 2016.

[101] Z. Kukelova, M. Bujnak, and T. Pajdla, “Automatic generator of minimal problem solvers,” *Computer Vision–ECCV 2008*, 2008.

[102] W. E. L. Grimson and T. Lozano-Perez, "Model-based recognition and localization from sparse range or tactile data," *The international journal of robotics research*, vol. 3, no. 3, pp. 3–35, 1984.

[103] S. Ramalingam, Y. Taguchi, T. K. Marks, and O. Tuzel, "P2Π: A minimal solution for registration of 3D points to 3D planes," in *European Conference on Computer Vision*. Springer, 2010, pp. 436–449.

[104] S. Ramalingam and Y. Taguchi, "A theory of minimal 3D point to 3D plane registration and its generalization," *International journal of computer vision*, 2013.

[105] A. Censi, "An ICP variant using a point-to-line metric," *Robotics and Automation, 2008. ICRA 2008. IEEE*, 2008.

[106] J. Saunderson, P. Parrilo, and A. S. Willsky, "Semidefinite relaxations for optimization problems over rotation matrices," in *Decision and Control (CDC), 2014 IEEE 53rd Annual Conference On*. IEEE, 2014, pp. 160–166.

[107] D. M. Rosen, C. DuHadway, and J. J. Leonard, "A convex relaxation for approximate global optimization in simultaneous localization and mapping," in *Robotics and Automation (ICRA), 2015 IEEE International Conference On*. IEEE, 2015, pp. 5822–5829.

[108] Y. Khoo and A. Kapoor, "Non-iterative rigid 2D/3D point-set registration using semidefinite programming," *IEEE Transactions on Image Processing*, vol. 25, no. 7, pp. 2956–2970, 2016.

[109] Y. Ding, "On efficient semidefinite relaxations for quadratically constrained quadratic programming," Ph.D. dissertation, University of Waterloo, 2007.

[110] L. Carlone, G. C. Calafiore, C. Tommolillo, and F. Dellaert, "Planar Pose Graph Optimization: Duality, Optimal Solutions, and Verification," *IEEE Transactions on Robotics*, vol. 32, no. 3, pp. 545–565, 2016.

[111] J. Briales and J. González-Jiménez, "Fast Global Optimality Verification in 3D SLAM," in *Proc. Int. Conf. Intell. Robots Syst. (IROS)*. Daejeon: IEEE/RSJ, 2016.

[112] D. M. Rosen, L. Carlone, A. S. Bandeira, and J. J. Leonard, "A Certifiably Correct Algorithm for Synchronization over the Special Euclidean Group," in *Int. Workshop Alg. Found. Robot. (WAFR)*, San Francisco, Nov. 2016.

[113] Y. Nesterov, H. Wolkowicz, and Y. Ye, "Semidefinite programming relaxations of nonconvex quadratic optimization," in *Handbook of Semidefinite Programming*. Springer, 2000, pp. 361–419.

[114] P. A. Parrilo and S. Lall, "Semidefinite programming relaxations and algebraic optimization in control," *European Journal of Control*, vol. 9, no. 2, pp. 307–321, 2003.

[115] J. P. Ruiz and I. E. Grossmann, “Using redundancy to strengthen the relaxation for the global optimization of MINLP problems,” *Computers & Chemical Engineering*, vol. 35, no. 12, pp. 2729–2740, 2011.

[116] K. Anstreicher and H. Wolkowicz, “On Lagrangian relaxation of quadratic matrix constraints,” *SIAM Journal on Matrix Analysis and Applications*, vol. 22, no. 1, pp. 41–55, 2000.

[117] G. C. Calafiore and L. El Ghaoui, *Optimization Models*. Cambridge university press, 2014.

[118] R. Tron, D. M. Rosen, and L. Carlone, “On the inclusion of determinant constraints in lagrangian duality for 3D SLAM,” in *Robotics: Science and Systems (RSS) in the Workshop “The Problem of Mobile Sensors*, 2015.

[119] M. Grant and S. Boyd, “{CVX}: Matlab Software for Disciplined Convex Programming, version 2.1,” 2014.

[120] P. L. Fackler, “Notes on Matrix Calculus,” North Carolina State University, Tech. Rep., 2005.

[121] B. Triggs, P. McLauchlan, R. Hartley, and A. Fitzgibbon, “Bundle Adjustment - A Modern Synthesis,” in *Proceedings of the International Workshop on Vision Algorithms: Theory and Practice (ICCV)*, Corfu, Greece, 1999, pp. 298–372.

[122] D. Nister and H. Stewenius, “Scalable recognition with a vocabulary tree,” in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’06)*, vol. 2. Ieee, 2006, pp. 2161–2168.

[123] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardós, “ORB-SLAM: A Versatile and Accurate Monocular SLAM System,” *Trans. Robot. (TRO)*, vol. 31, no. 5, pp. 1147–1163, 2015.

[124] E. Kruppa, “Zur {E}rmittlung eines {O}bjektes aus zwei {P}erspektiven mit innerer {O}rientierung,” *Sitzgsber. Akad. Wien, Math. Naturw. Abt., IIa.*, vol. 122, pp. 1939–1948, 1913.

[125] D. Nistér, “An Efficient Solution to the Five-Point Relative Pose Problem,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 26, no. 6, pp. 756–777, 2004.

[126] H. Stewenius, C. Engels, and D. Nister, “Recent Developments on Direct Relative Orientation,” *{ISPRS} Journal of Photogrammetry and Remote Sensing*, vol. 60, no. 4, pp. 284–294, 2006.

[127] L. Kneip and S. Lynen, “Direct optimization of frame-to-frame rotation,” in *Computer Vision (ICCV), 2013 IEEE International Conference On*. IEEE, 2013, pp. 2352–2359.

[128] L. Kneip, H. Li, and Y. Seo, “Upnp: An optimal o (n) solution to the absolute pose problem with universal applicability,” in *Computer Vision–ECCV 2014*. Springer, 2014, pp. 127–142.

[129] H. C. Longuet-Higgins, "A computer algorithm for reconstructing a scene from two projections," *Nature*, vol. 293, no. 5828, pp. 133–135, 1981.

[130] R. I. Hartley, "{I}n {D}efense of the {E}ight-{P}oint {A}lgorithm," *TPAMI*, vol. 19, no. 6, pp. 580–593, 1997.

[131] R. I. Hartley and F. Kahl, "Global optimization through rotation space search," *International Journal of Computer Vision*, vol. 82, no. 1, pp. 64–79, 2009.

[132] J. Yang, H. Li, and Y. Jia, "Optimal Essential Matrix Estimation via Inlier-Set Maximisation," in *ECCV*, 2014.

[133] D. M. Rosen, C. DuHadway, and J. J. Leonard, "A convex relaxation for approximate global optimization in simultaneous localization and mapping," in *Proc. Int. Conf. Robot. and Autom. (ICRA)*. Seattle: IEEE, 2015, pp. 5822–5829.

[134] N. Boumal, "A Riemannian low-rank method for optimization over semidefinite matrices with block-diagonal constraints," *arXiv preprint arXiv:1506.00575*, 2015.

[135] A. Bandeira, N. Boumal, and A. Singer, "Tightness of the maximum likelihood semidefinite relaxation for angular synchronization," *Mathematical Programming*, 2016.

[136] L. Kneip, R. Siegwart, and M. Pollefeys, *Finding the Exact Rotation between Two Images Independently of the Translation*. Springer, 2012.

[137] R. Tron, D. M. Rosen, and L. Carlone, "On the Inclusion of Determinant Constraints in Lagrangian Duality for 3D SLAM," in *Proc. Robotics: Science and Systems (RSS), Workshop "The Problem of Mobile Sensors"*, 2015.

[138] N. Z. Shor, "Quadratic optimization problems," *Soviet Journal of Computer and Systems Sciences*, vol. 25, no. 6, pp. 1–11, 1987.

[139] L. Vandenberghe and S. Boyd, "Semidefinite programming," *SIAM review*, vol. 38, no. 1, pp. 49–95, 1996.

[140] H. Wolkowicz, "A note on lack of strong duality for quadratic problems with orthogonal constraints," *European Journal of Operational Research*, vol. 143, no. 2, pp. 356–364, 2002.

[141] Z.-Q. Luo, W.-K. Ma, A. M.-C. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal Processing Magazine*, vol. 27, no. 3, pp. 20–34, 2010.

[142] J. F. Sturm, "Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones," *Optimization methods and software*, vol. 11, no. 1-4, pp. 625–653, 1999.

[143] K.-C. Toh, M. J. Todd, and R. H. Tütüncü, "SDPT3—a MATLAB software package for semidefinite programming, version 1.3," *Optimization methods and software*, vol. 11, no. 1-4, pp. 545–581, 1999.

[144] A. Bandeira, “A note on probably certifiably correct algorithms,” *Comptes Rendus Mathematique*, 2016.

[145] D. Henrion, J.-B. Lasserre, and J. Löfberg, “GloptiPoly 3: Moments, optimization and semidefinite programming,” *Optimization Methods & Software*, vol. 24, no. 4-5, pp. 761–779, 2009.

[146] K. Schäcke, “On the kronecker product,” University of Waterloo, Tech. Rep., 2013.

[147] M. Kaess, H. Johannsson, R. Roberts, V. Ila, J. J. Leonard, and F. Dellaert, “iSAM2: Incremental smoothing and mapping using the Bayes tree,” *Int. J. Robot. Res. (IJRR)*, vol. 31, no. 2, pp. 216–235, 2011.

[148] D. M. Rosen, M. Kaess, and J. J. Leonard, “RISE: An incremental trust-region method for robust online sparse least-squares estimation,” *Trans. Robot. (TRO)*, vol. 30, no. 5, pp. 1091–1108, 2014.

[149] L. Carlone, R. Aragues, J. A. Castellanos, and B. Bona, “A fast and accurate approximation for planar pose graph optimization,” *The International Journal of Robotics Research*, 2014.

[150] L. Carlone and A. Censi, “From Angular Manifolds to the Integer Lattice: Guaranteed Orientation Estimation With Application to Pose Graph Optimization,” *Trans. Robot. (TRO)*, vol. 30, no. 2, pp. 475–492, Apr. 2014.

[151] L. Carlone, R. Tron, K. Daniilidis, and F. Dellaert, “Initialization techniques for 3D SLAM: A survey on rotation estimation and its use in pose graph optimization,” in *Robotics and Automation (ICRA), 2015 IEEE International Conference On*. IEEE, 2015, pp. 4597–4604.

[152] E. Olson, J. Leonard, and S. Teller, “Fast iterative alignment of pose graphs with poor initial estimates,” in *Proc. Int. Conf. Robot. and Autom. (ICRA)*. IEEE, 2006, pp. 2262–2269.

[153] R. Tron, B. Afsari, and R. Vidal, “Intrinsic consensus on SO (3) with almost-global convergence.” in *CDC*, 2012, pp. 2052–2058.

[154] S. Huang, Y. Lai, U. Frese, and G. Dissanayake, “How far is SLAM from a linear least squares problem?” in *Proc. Int. Conf. Intell. Robots Syst. (IROS)*. IEEE, 2010, pp. 3011–3016.

[155] H. Wang, G. Hu, S. Huang, and G. Dissanayake, “On the Structure of Nonlinearities in Pose Graph SLAM.” in *Robotics: Science and Systems*, 2012.

[156] H. Wang, S. Huang, U. Frese, and G. Dissanayake, “The nonlinearity structure of point feature SLAM problems with spherical covariance matrices,” *Automatica*, vol. 49, no. 10, pp. 3112–3119, 2013.

[157] L. Carlone, “A convergence analysis for pose graph optimization via gauss-newton methods,” in *Robotics and Automation (ICRA), 2013 IEEE International Conference On*. IEEE, 2013, pp. 965–972.

[158] K. Khosoussi, S. Huang, and G. Dissanayake, “Novel insights into the impact of graph structure on SLAM,” in *Proc. Int. Conf. Intell. Robots Syst. (IROS)*. IEEE, 2014, pp. 2707–2714.

[159] R. Sim and N. Roy, “Global a-optimal robot exploration in slam,” in *Proceedings of the 2005 IEEE International Conference on Robotics and Automation*. IEEE, 2005, pp. 661–666.

[160] L. Carlone and F. Dellaert, “Duality-based verification techniques for 2D SLAM,” in *Intl. Conf. on Robotics and Automation (ICRA)*, 2015.

[161] K. V. Mardia and P. E. Jupp, *Directional Statistics*. John Wiley & Sons, 2009, vol. 494.

[162] J. Sturm, N. Engelhard, F. Endres, W. Burgard, and D. Cremers, “A Benchmark for the Evaluation of RGB-D SLAM Systems,” in *Proc. Int. Conf. Intell. Robots Syst. (IROS)*, 2012.

[163] D. Martinec and T. Pajdla, “Robust rotation and translation estimation in multiview reconstruction,” in *Computer Vision and Pattern Recognition, 2007. CVPR’07. IEEE Conference On*. IEEE, 2007, pp. 1–8.

[164] M. Arie-Nachimson, S. Z. Kovalsky, I. Kemelmacher-Shlizerman, A. Singer, and R. Basri, “Global motion estimation from point matches,” in *3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT), 2012 Second International Conference On*. IEEE, 2012, pp. 81–88.

[165] G. Piovan, I. Shames, B. Fidan, F. Bullo, and B. D. O. Anderson, “On frame and orientation localization for relative sensing networks,” *Automatica*, vol. 49, no. 1, pp. 206–213, 2013.

[166] J. Briales and J. Gonzalez-Jimenez, “Initialization of 3D Pose Graph Optimization using Lagrangian duality - Supplementary Material.”

[167] N. Boumal, A. Singer, and P.-A. Absil, “Robust estimation of rotations from relative measurements by maximum likelihood,” in *Decision and Control (CDC), 2013 IEEE 52nd Annual Conference On*, 2013, pp. 1156–1161.

[168] O. Ozyesil, N. Sharon, and A. Singer, “Synchronization over Cartan motion groups via contraction,” *arXiv preprint arXiv:1612.00059*, Nov. 2016.

[169] L. Carlone, G. C. Calafiore, C. Tommolillo, and F. Dellaert, “Planar Pose Graph Optimization: Duality, Optimal Solutions, and Verification,” *Trans. Robot. (TRO)*, vol. 32, no. 3, pp. 545–565, 2016.

[170] A. Chiuso, G. Picci, S. Soatto *et al.*, “Wide-sense estimation on the special orthogonal group,” *Communications in Information & Systems*, vol. 8, no. 3, pp. 185–200, 2008.

[171] R. Tron, X. Zhou, and K. Daniilidis, “A Survey on Rotation Optimization in Structure From Motion,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2016, pp. 77–85.

[172] N. Boumal, B. Mishra, P.-A. Absil, and R. Sepulchre, “Manopt, a Matlab toolbox for optimization on manifolds,” *Journal of Machine Learning Research (JMLR)*, vol. 15, no. 1, pp. 1455–1459, 2014.

[173] C. Kleiber and A. Zeileis, “Visualizing Count Data Regressions Using Rootograms,” *http://arxiv.org/abs/1605.01311*, May 2016.

[174] N. Boumal, V. Voroninski, and A. Bandeira, “The non-convex Burer-Monteiro approach works on smooth semidefinite programs,” in *Adv. Neur. Inf. Proc. Sys. (NIPS)*, 2016, pp. 2757–2765.

[175] A. Singer and H.-T. Wu, “Vector diffusion maps and the connection Laplacian,” *Communications on pure and applied mathematics*, vol. 65, no. 8, pp. 1067–1144, 2012.

[176] J. Briales and J. Gonzalez-Jimenez, “Cartan-Sync: Fast and Global SE(d)-Synchronization - Supplementary Material.”

[177] A. S. Bandeira, A. Singer, and D. A. Spielman, “A Cheeger inequality for the graph connection Laplacian,” *SIAM Journal on Matrix Analysis and Applications*, vol. 34, no. 4, pp. 1611–1630, 2013.

[178] S. Burer and R. D. C. Monteiro, “A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization,” *Mathematical Programming*, vol. 95, no. 2, pp. 329–357, 2003.

[179] ——, “Local minima and convergence in low-rank semidefinite programming,” *Mathematical Programming*, vol. 103, no. 3, pp. 427–444, 2005.

[180] A. Edelman, T. A. Arias, and S. T. Smith, “The geometry of algorithms with orthogonality constraints,” *SIAM Journal on Matrix Analysis and Applications*, vol. 20, no. 2, pp. 303–353, 1998.

[181] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, 2009.

[182] P.-A. Absil, C. G. Baker, and K. A. Gallivan, “Trust-region methods on Riemannian manifolds,” *Foundations of Computational Mathematics*, vol. 7, no. 3, pp. 303–330, 2007.

[183] A. R. Conn, N. I. M. Gould, and P. L. Toint, *Trust Region Methods*. Philadelphia, USA: Society for Industrial and Applied Mathematics (SIAM), 2000.

[184] Y. Latif, C. Cadena, and J. Neira, “Robust graph SLAM back-ends: A comparative analysis,” in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Sep. 2014, pp. 2683–2690.